Theoretical foundations of environmental pollution monitoring

Andrii Biloshchytskyi[1,2†], Olexander Kuchanskiy[1,*,†], Alexander Neftissov[1,†], Svitlana Biloshchytska [1,2†] and Sapar Toxanov [1,†]

[1] *Astana IT University, Astana 010000, Kazakhstan*
[2] *Kyiv National University of Construction and Architecture, Kyiv 03037, Ukraine*

Abstract
The article discusses the basic concepts and features of environmental monitoring. The necessity of improving the effectiveness of monitoring and the main approaches to solving them by improving methods and technologies is substantiated. Analysis of the properties of time series of pollutants shows that they can be classified into three classes: substances with a pronounced seasonal component, substances with a pronounced trend and random variables. The problem of environmental monitoring is formalized in two formulations: point-based and planar. The main stages of environmental monitoring are highlighted. Fundamental differences and new trends in the use of innovative technologies for monitoring environmental pollution parameters have been identified. A scientific hypothesis has been formulated that defines the author's vision of the organization of environmental monitoring from the point of view of combining software and hardware complexes and using trend models to predict environmental pollution parameters. By formalizing the problem of environmental monitoring, the structure of an information system for environmental monitoring is proposed. It is indicated that the construction of an air pollution monitoring system is also important for the holistic and safe operation of some critical infrastructure facilities, including power plants, processing and chemical plants, airports, tunnels and subways, etc. In the case of poor-quality measurements of the environment near or inside these facilities, irreparable consequences for the environment, health and life of many people may occur.

Keywords
Environmental monitoring, emissions into the environment, pollution forecasting methods, information and analytical systems.

CEUR-WS.org/Vol-3966/W2Paper7.pdf

1. Introduction

Ensuring a balanced solution to the tasks of preserving a favorable environment, applying new approaches to environmental protection and observing the economic interests of both enterprises and the entire population requires a focused scientific approach. In recent years, there has been a close relationship between economic development and changes in the environment, and the mutual influence of the state of the environment on economic development and the results of economic activity on the state of the environment is growing.

In the face of a constantly deteriorating environmental situation, the scientific basis for managing anthropogenic impact, multifactorial analysis of pollution formation, combined with an operational forecast of pollution levels, is the only effective way to solve the problem.

Environmental pollution research includes the study of air pollution, groundwater and surface water pollution, soil pollution, and impact on the biosphere. Each type of pollution requires its models and research methods, as well as forecasting.

DTESI 2024: 9th International Conference on Digital Technologies in Education, Science and Industry, October 16–17, 2024, Almaty, Kazakhstan
* Corresponding author.
† These authors contributed equally.
A.B@astanait.edu.kz (A. Biloshchytskyi); a.kuchanskyi@astanait.edu.kz (O. Kuchanskiy); alexandr.neftissov@astanait.edu.kz (A. Neftissov)
0000-0001-9548-1959 (A. Biloshchytskyi); 0000-0003-1277-8031 (O. Kuchanskiy); 0000-0003-4079-2025 (A. Neftissov)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

A meta-analysis of sources shows a significant increase in the international community's interest in studying environmental pollution. Most research publications on environmental pollution were made after 2012 [1, 2]. The primary purpose of these studies is to develop new methods for predicting the state of pollution, studying environmental monitoring systems and creating models of dependencies between pollution factors. At the same time, 60% of publications are devoted to forecasting, confirming this research area's prospects.

Fig. 1 shows the graph of changes in the number of scientific publications devoted to the study of environmental pollution found in the online version of the Science Citation Index (SCI-Expanded) from 1991 to 2017. The keywords used for the search were: ‘‘pollution’’, ‘‘pollutions’’, ‘‘polluted’’, ‘‘polluting’’, ‘‘pollutant’’, ‘‘pollutants’’, ‘‘pollute’’, ‘‘pollutes’’, ‘‘contamination’’, ‘‘contaminations’’, ‘‘contaminate’’, ‘‘contaminant’’, ‘‘contaminants’’, ‘‘contaminated’’, ‘‘contaminating’’, ‘‘estuary’’, ‘‘estuaries’’, ‘‘estuarius’’, ‘‘estuaria’’, ‘‘estuaries’’, ‘‘estuarial’’, ‘‘estuarian’’, and ‘‘estuarine’’.

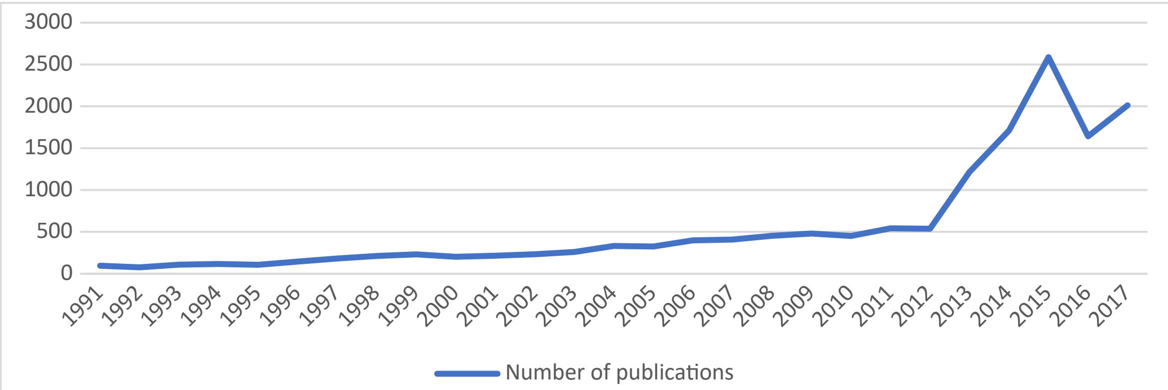

Figure 1: Graph of changes in the number of scientific publications devoted to the study of environmental pollution [2].

The importance of environmental protection research is also confirmed by the fact that governments of all leading countries spend an average of 0.8% of their budgets (more than $600 billion) on environmental protection measures [3]. Among these expenditures, R&D ranks third.

Environmental monitoring is systematically collecting, analyzing and evaluating data on pollution levels, air, water, soil quality, and other factors that may affect environmental sustainability and the health of people and ecosystems. Environmental monitoring consists of the following general stages:

1. Data collection. This may include the installation of sensors and instruments to measure various parameters, such as the level of dissolved oxygen in water, the concentration of heavy metals in soil, sound levels, and others.
2. Data analysis. Evaluation of the data obtained and determination of the state of the environment, identification of possible sources of pollution and their impact on the environment.
3. Reporting and informing. Preparation of reports on the results of monitoring and dissemination of information on the state of the environment to stakeholders, including government agencies, NGOs and citizens.
4. Action planning. Based on the monitoring results, develop strategies and measures to reduce pollution and maintain environmental sustainability.

Forecasting time series of pollution parameters is necessary for high-quality monitoring of the environment for the following reasons:

1. Early detection of trends. Predicting pollution dynamics allows to detect trends and identify possible changes in the state of the environment over time.
2. Planning of measures. Forecast data can be used to develop strategies and measures to improve the environment and reduce pollution.

3. Monitoring the effectiveness of measures. Comparison of predicted data with actual data allows to evaluate the effectiveness of the measures taken and adjust strategies as necessary.
4. Prevention of crises. Forecasting can help identify potential crises and take measures to prevent or minimize their consequences.

Thus, air quality monitoring methods and models are essential for various stakeholders interested in environmental protection, public health, and effective solutions to air pollution problems. Forecasting is a necessary element of the monitoring system, but it allows to identify undesirable trends in the environment in advance and correct them.

2. Research methods

The task of monitoring environmental pollution Environmental forecasting has three main classes: expert methods, modeling, and extrapolation methods [4]. In cases where the data cannot be formalized and structured, it is relevant to use expert forecasting methods. Expert forecasting methods for environmental pollution parameters involve qualified experts to assess and predict the pollution level. Experts can use their knowledge and experience to assess the impact of various factors on the environment and develop forecasts. The Delphi method can also be used in this case. This expert forecasting procedure involves an iterative process of interviewing a group of experts. Experts make their forecasts and then analyze and discuss the results to obtain agreed forecasts. The Mind Maps method is also used for this task. This graphical method allows experts to visualize and systematize information about various factors affecting environmental pollution and their possible consequences.

Using historical data or data from similar situations to predict future pollution levels is also a critical approach. In addition, the development of various scenarios is often used to predict possible levels of environmental pollution depending on various conditions and factors.

Extrapolation methods are most often used for short-term forecasts. These methods are based on the study of data, their quantitative and qualitative analysis for previous periods. In cases where the environmental situation is not subject to sharp changes, trends in the situation's dynamics for the next forecast period are determined. Recently, modeling methods using computer technology have become the most widely used.

There are three main approaches to forecasting the state of environmental pollution:

1. Works [5, 6, 7] use an approach based on pattern recognition using neural networks.
2. The possibility of using methods based on regression analysis is shown in [8, 9, 10, 11].
3. The authors in [12-16] apply time series analysis methods, particularly trend forecasting methods.

The use of neural networks for environmental forecasting has a long history. In [5], five models of neural networks (NN), a linear statistical model, and a deterministic modeling system (DET) were compared to predict NO2 and PM10 concentrations in urban areas. The time series of NO2 and PM10 concentrations measured at two stations in the center of Helsinki from 1996 to 1999 on an hourly basis were considered. The data set required preliminary processing. Missing values were replaced to obtain a harmonized database. Comparisons were made using three criteria: the index of agreement (IA), the quadratic correlation coefficient (R2), and the fractional offset. The results obtained with different nonlinear NN models agree with the measured NO2 concentration data. In the case of NO2, the nonlinear NN models predict the crown concentration slightly better than DET. NN models perform better than the statistical linear model for predicting NO2 and PM10 concentrations. In the case of PM10, NN models were not as good as for NO2.

3. The task of monitoring environmental pollution

Paper [6] shows that modeling real-world processes, such as air quality, is a challenging task, both because of the chaotic and nonlinear nature of the phenomenon and because of the high dimensionality of the samples. Although neural networks have been successfully used in this area, the choice of network architecture still needs to be improved and more time-consuming when developing a model for a practical situation. The study proposes to use a parallel genetic algorithm (GA) for selecting input data and developing the architecture of a multi-layer perceptron model to predict nitrogen dioxide concentration at a high-traffic urban transport station in Helsinki. The results showed that the genetic algorithm is a suitable tool for solving practical problems of neural network design. However, it was noted that the evaluation of NN models is a computationally complex process, which sets limits for the application of this method. The authors also needed help tuning the GA parameters for the problem under consideration.

Paper [7] aims to compare two fundamentally different forecasting methods using a neural network. They are evaluated in terms of regression with periodic scalars. Self-organizing maps (SOM) are a form of competitive learning in which a neural network learns the data structure. It is shown that Multi-layer perceptrons (MLPs) are capable of learning complex relationships between input and output variables. In addition, the positive impact of removing periodic components on the quality of neural network training is shown. The methods were evaluated using a time series of NO2 concentrations. The estimated values for forecasting were calculated in three ways:

- using only periodic components;
- applying neural network methods to the residual values after removing periodic components;
- applying only the output data to the neural networks

The results showed that the best forecast predictions can be achieved by combining the periodic regression method and neural algorithms. However, the advantage of directly applying the MLP network to the raw data is not significant.

Paper [13] discusses the BFAST (Breaks For Additive Seasonal & Trend) method. This method combines methods for detecting changes in the behavior of time series with methods for decomposing series into components that determine trend changes, seasonal changes, and random components.

According to this method, the time series model looks like this:

$$Y_t = T_t + S_t + e_t, \tag{1}$$

where Y_t is the time series data recorded at time t;

T_t – trend component;

S_t is seasonal component;

e_t are residual, random components $t=\overline{1,n}$,, n is number of observations or number of elements in the image time series.

The residual components represent variations in the time series that characterize random deviations from the trend or seasonal components. In this model, the trend component is assumed to be piecewise linear, which means that it is specified in the form:

$$T_t = a_i + t \cdot b_I, \tag{2}$$

where $r_{i-1} < t \le r_I$, $i=\overline{1,m}$ – control points of observation.

To determine the seasonal component, you can set a linear harmonic regression model:

$$S_t=\sum_{k=1}^{K}\left(\gamma_{jk}\sin\left(\frac{2\pi kt}{\lambda}\right)+\chi_{jk}\cos\left(\frac{2\pi kt}{\lambda}\right)\right), \tag{3}$$

where $\gamma_{jk}=\alpha_{jk}\cos\beta_{jk}, \quad \chi_{jk}=\alpha_{jk}\sin\beta_{jk}$ are model coefficients.

The amplitude can be defined as

$$A_{jk}=\sqrt{\gamma_{jk}^2+\chi_{jk}^2}, \tag{4}$$

and the phase for the frequency $\frac{\lambda}{k}$ is defined as

$$\beta_{jk}=\frac{1}{tg\left(\frac{\chi_{jk}}{\gamma_{jk}}\right)}. \tag{5}$$

The described model has the following advantages over the conventional seasonal model:

1. The model is less sensitive to short-term changes and noise.
2. A few observations are not required to calculate the parameters of the multiple regression model.

Applied models of geostatistical analysis were studied in [14]. The further development of these studies was the work [15], which investigated approaches to geostatistical modeling using variogram models. Studies on multivariate analysis, which allow for the selection of analysis options, are presented in [16]. Applied work on using geostatistical methods in the study of the environment and environmental problems is described in [17]. Most of the described methods are designed to work with continuous distributions of geostatistical indicators in the environment, so the mechanisms for processing discrete values need further improvement.

The peculiarity of air pollution is its ability to spread pollutants over vast distances and its significant dependence on weather conditions. Also, the atmosphere should be considered not only as a polluted environment but also as a mediator of anthropogenic pollution of other components of nature. The problem of anthropogenic and technogenic pollution is especially relevant in large cities with many industrial enterprises, vehicles and populations. Works [7, 18] analyzed the time series of 16 air pollutants and investigated their trend, seasonal, and random components (Table 1).

According to the identified patterns in the dynamic series, pollutants can be classified into three classes:

1. Substances with a pronounced seasonal component: benzopyrene, sulfur dioxide, carbon monoxide. This cycle is because in winter, the emissions of these pollutants from thermal power plants and motor vehicles increase significantly. Summer and winter periods affect the concentration of these pollutants in the air;
2. Substances with a pronounced trend: benzene, toluene, ethylbenzene, nitrogen oxide. In addition to the seasonal component, these pollutants have a pronounced upward trend in concentration.
3. Random values in which it is difficult to identify the seasonal component: trichloromethane, ammonia. Their level is influenced by random events (non-periodic processes, volley and accidental emissions, unfavorable meteorological conditions, etc.)

The use of trend models is possible only for forecasting the pollution level of substances of the first and second groups In addition, the study shows that the contribution of the random component to the structure of the time series of each group is large. This means that there are many hidden factors.

Table 1
The time series of 16 air pollutants and investigated their trend, seasonal, and random components

№	Pollutant	Contribution of the component		
1	Dust	0,0051	0,4891	0,5059
2	Sulfur dioxide	0,0435	0,4343	0,5222
3	Carbon monoxide	0,0378	0,4718	0,4903
4	Nitrogen dioxide	0,0193	0,4831	0,4976
5	Nitrogen oxide	0,0987	0,3987	0,5026
6	Hydrogen sulfide	0,0538	0,4481	0,4981
7	Phenol	0,0312	0,4589	0,5099
8	Hydrogen chloride	0,0430	0,4670	0,4900
9	Ammonia	0,0082	0,1828	0,8091
10	Formaldehyde	0,0796	0,4529	0,4675
11	Benzene	0,3368	0,3110	0,3523
12	Toluene	0,2585	0,3421	0,3995
13	Ethyl benzene	0,0907	0,3819	0,5274
14	Dust	0,0659	0,1296	0,8045
15	Sulfur dioxide	0,0237	0,4878	0,4886
16	Carbon monoxide	0,0194	0,4352	0,5454

Given the above, we can assume that using neural network-based models in air pollution forecasting tasks is a good option. Neural networks can consider hidden dependencies. Dynamic series form the basis for forming samples for training and testing neural networks.

4. Peculiarities of building systems for monitoring pollution of the scientific environment for environmental safety management

To obtain information on the dynamics of the content of harmful substances in the environment and to draw up maps of its pollution based on experimental data, it is necessary to measure the concentrations of pollutants in the air regularly. An automated information monitoring system (AIMS) is a system with a distributed organization of collection, processing, documentation and analysis of environmental parameters. In any environmental monitoring system, an AISM is an essential element and is designed to collect, process, and store information quickly and over the long term, forecast the state of the environment based on it, and provide information to local information centers, the management of enterprises and their environmental protection departments, and other information users. AISM provides the following functions:

- automatic measurement of monitored parameters;
- collection of information and its primary processing;
- control of deviations of current values of these parameters from their reference levels;
- display of information and formation of the operational situation;
- documentation of information;
- forecasting changes in the environment;
- transfer of information to interested parties and adjacent systems.

An automated control system (ACS) called Ecoinspector has been introduced in Ukraine [19]. This ACS is a comprehensive solution that includes hardware and software that can be divided into three parts:

- software for mobile devices and sensors for monitoring the parameters of the built environment, which performs the functions of registering information and taking samples and measurements performed directly at the monitoring site;
- server software that performs data storage and processing functions;
- software for a regular personal computer for other operations.

The system is based on a set of subsystems for processing data from one analytical department of the regional and national environmental inspectorate. The national-level software additionally has a set of subsystems for importing data from all analytical departments into a single database, as well as for processing them and generating various reports.

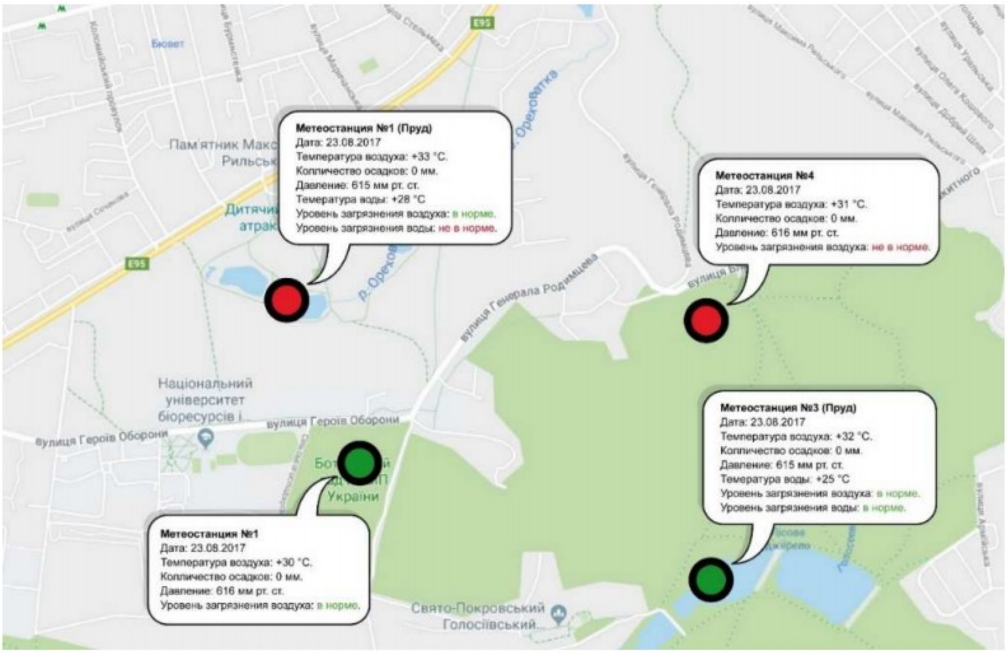

Figure 2: An interactive map of environmental monitoring on the example of the Inspector Meteo system.

The results of the system can be seen in real-time using an interactive map of environmental monitoring [20], which is available on the website of the Ministry of Ecology and Natural Resources of Ukraine (Fig. 3).

Similar environmental monitoring systems operate around the world, including the air pollution monitoring system in China (Fig. 4), which became the basis for the international project The World Air Quality Index [21].

The traditional approach to environmental monitoring involves observation points and centralized data processing. This approach is only sometimes economically feasible. Distributed networks are a concept in which individuals, groups, and communities are actively involved in collecting data to build a knowledge base. This is mainly done in two ways: using an extensive sensor network and using available devices (e.g., mobile phones) to create ad hoc networks.

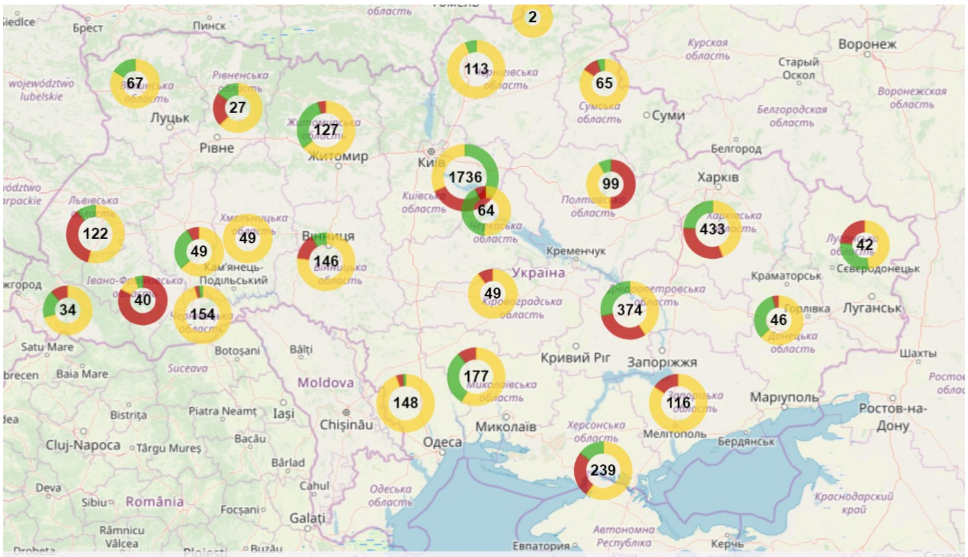

Figure 3: Interactive map of monitoring environment of the environment.

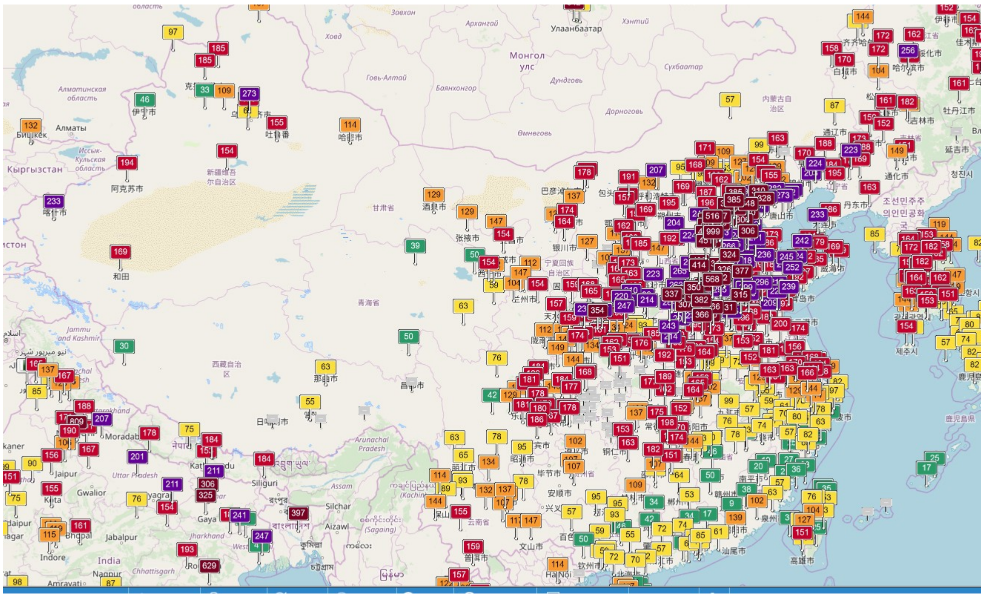

Figure 4: Air Pollution in China: Real-time Air Quality Index Visual Map.

Low-cost sensor technologies have the potential to revolutionize the field of air pollution monitoring by providing high-density air pollution data. Such data can complement traditional pollution monitoring, improve impact assessments, and raise community awareness of air pollution. However, data quality remains a significant challenge that hinders the widespread adoption of low-cost sensor technologies. Unreliable data can mislead users and potentially lead to alarming consequences, such as reporting acceptable levels of air pollutants when they exceed limits recognized as safe for human health [22, 23]. Paper [24] addresses the efficient deployment of low-cost sensors while ensuring sufficient data quality. For large sensor networks, where conventional calibration checks are impractical, statistical methods of data quality assurance should be used. There is a need to develop mathematical and statistical methods for sensor calibration, fault detection, and data quality assurance.

Water monitoring is a much more expensive and technologically complex process. Thus, a bioelectronic nose was described for real-time water quality assessment [25]. The nose is built on the principle of a human olfactory receptor based on a single-walled carbon nanotube field-effect transistor (swCNT-FET). The bioelectronic nose can selectively detect Geosmin (GSM) and 2-methylisoborneol (MIB) in low concentrations. The main problem of this sensor is the need to use a carbon nanotube field-effect transistor.

A technical report on water monitoring tools was developed as part of the Water Framework Directive (WFD), 2000/60/EC study [26-30]. It identifies potential exposure-based tools (e.g., biomarkers and bioassays) that can be used in different monitoring programs (surveillance, operational and investigative) that link chemical and ecological status assessment.

Let's consider the problem of estimating environmental pollution in two formulations: point and plane. Let's assume it is necessary to estimate environmental pollution at a certain point. A set of indicators can estimate environmental pollution. Let

$$R=\left(r_1,r_2,\ldots,r_n\right), \tag{6}$$

is a vector of real numbers describing the state of the environment, where n is the number of indicators. Each vector coordinate is a specific indicator, for example, the concentration of sulfur dioxide or carbon monoxide in the air, the concentration of nitrates in water, etc. The relevant indicators can be obtained both with the help of appropriate technical means (weather stations, mobile and stationary sensors, etc.) and with the help of services.

The state of the environment is not a stationary value and changes over time. Therefore, environmental indicators should be considered as time-dependent functions. That is.

$$R(t)=\left(r_1(t),r_2(t),\ldots,r_n(t)\right), \tag{7}$$

where t is a certain time. For the sake of simplicity, we will assume that the indicators are updated with a certain period (hourly, daily, monthly). Then, without limiting the generality, we will consider time as a discrete value. That is

$$t_i=t_0+\Delta t\circ i, \tag{8}$$

where t_0 is the initial moment of time from which the environmental state is observed, Δt is the frequency of observation, and a $i=\overline{1,m}$, where m is the number of observations.

Let's define $r_j\left(t_i\right)$as r_j^i. Then

$$R\left(t_i\right)=\left(r_1\left(t_i\right),r_2\left(t_i\right),\ldots,r_n\left(t_i\right)\right)=\left(r_1^i,r_2^i,\ldots,r_n^i\right). \tag{9}$$

Then the task of assessing environmental pollution can be divided into the following stages:

1. Collecting data on the history of environmental pollution.
2. Observation of the current state of the environment.
3. Forecasting the state of environmental pollution in the future.

To solve the first task, building a database that stores the history of environmental pollution is necessary. There are two possible ways of storing it. The first way is to store the history of the state of the environment as a set of dynamic series, each of which reflects the change in one indicator. The second way is to save a sequence of vectors, each of which reflects the state of the environment at a certain point in time.

The second task requires a data source, a data transmission channel, and methods for converting information. The source of environmental data can be either hardware or other environmental monitoring services. The data transmission channel depends on the data source. Most often, the transmission channel is the Internet, but sometimes, it is necessary to transmit data through service protocols, such as Zigbee [31-36], to a form in which it can be stored in the system described in the first task.

Consider the third problem for the case when only one indicator needs to be forecasted. Then the forecasting task is to calculate the values of the pollutant indicator with a horizon θ>1, i.e. for each time point m+1, m+2, ... m+ θ. In other words, it is necessary to continue the dynamic series of pollution indicators:

$$R^{¿}=\left(\bar{r}_{n+1},\bar{r}_{n+2},\ldots,\bar{r}_{n+\theta}\right), \tag{10}$$

where the horizon θ is fixed before the forecast is calculated.

Let p be the size of the retrospective sample, i.e., the size of the area of the time series immediately following the point at which the forecast is calculated (point t_m), and which is involved in calculating the forecast values for p<m. The functional relationship based on which the values are predicted is called a forecasting model. Moreover, $\bar{r}_{n+\tau}$is the predicted estimate calculated at point r_n for τ points ahead with period $\tau=\overline{1,\theta}$. If we formally denote such a model as f, then the forecast calculated at point r_n for one point ahead or with a period of 1 can be defined as follows: $\bar{r}_{n+1}=f\left(r_{n-m+1},r_{n-m},\ldots,r_n\right)$.

As shown earlier, various forecasting models can be used for forecasting: regression, trend, neural network, etc. It was also shown that different models should be used for different environments of environmental indicators. Therefore, an important task is to build a method that takes into account a priori and a posteriori information and allows to improve the quality of the forecast by choosing a forecasting model that is better suited to a particular case.

The problem of assessing environmental pollution in a plane setting has much in common with a point setting. Similarly, the task consists of three stages: collection, observation, and forecasting.

The key difference in this setting is the presence of a whole observation network. Then the information about the state of the environment can be described as a set of tuples $\left\langle R_i,C_l\right\rangle$, where $R_i$is a vector reflecting the state of environmental pollution indicators at time t_i, and 〖C〗_l is information about the location where the relevant data were obtained. They are set in a specific coordinate system. It should also be noted that a significant part of the methods of forecasting and searching for the existence of a relationship between the greatnesses on the plane are based on the assumption that the coordinates are set in the Cartesian system. Observations of the state of the environment are linked to geographic coordinates.

The geographic coordinate system is used to determine the position of points on the earth's surface relative to the equator and the initial (zero) meridian. The coordinates are angular quantities: geographic latitude B and geographic longitude L. Longitude (the angle between the meridian plane at the point of observation and the zero (Greenwich) meridian), latitude (the angle between the straight line and the equator plane) determine the position of the point on the Earth's surface. Measured in degrees (°), longitude is from 0° to 180° west and east of Greenwich, latitude is from 0° to 90° north, from 0° to -90° south of the equator.

The geographic coordinate system is spherical. Therefore, a conversion formula should be used to convert to the Cartesian system. Considering all the above, the information system should include the following subsystems [37-42]:

1. Subsystems for collecting information about the state of the environment. This subsystem includes hardware for measuring environmental indicators, APIs for importing from other environmental monitoring systems, and methods for converting data to a single format used in the data storage subsystem.
2. Data storage and accumulation should be optimized considering the specifics of the data to be stored.
3. The environmental forecasting subsystem includes forecasting models and methods for selecting which model should be used in a particular case to achieve greater forecasting accuracy.
4. The user interaction subsystem is one of the most essential parts of the information system. It should present information in a convenient form. In particular, the presentation of reports, interactive maps of the state of the environment, and recommendations on dangerous

changes in environmental factors, such as exceeding the maximum permissible concentrations of certain pollutants.

Each subsystem can be considered as a separate module. The system's modular structure will allow you to expand and modify the capabilities of each module independently of the others. The modular structure also increases the stability and flexibility of the system. Given the modern approach to software development, the modular approach allows you to implement a microservice approach when the system consists of a set of independent microservices.

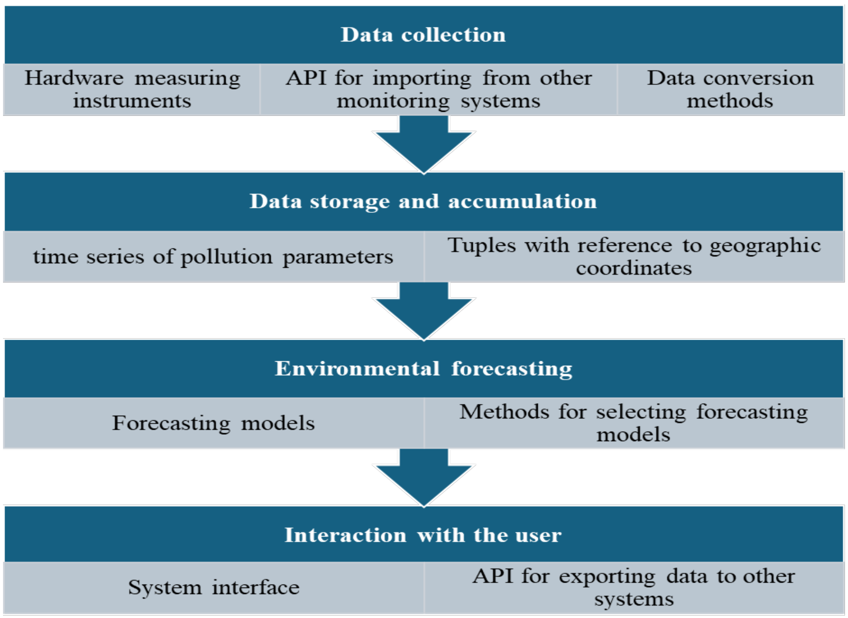

Figure 5: Conceptual diagram of the environmental monitoring system.

Further studies will consider the implementation of the proposed methods and an information and analytical system for monitoring emissions into the environment, as well as what consequences this entails.

Acknowledgment. This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number BR21882258"Development of Intelligent Information and Communication Systems Complex for Environmental Emission Monitoring to Make Decisions on Carbon Neutrality".

5. Conclusions

The basic concepts and features of environmental monitoring are considered. The necessity of increasing monitoring efficiency and the main approaches to their solution by improving methods and technologies are substantiated. The analysis of the properties of time series of pollutants shows that they can be classified into three classes: substances with a pronounced seasonal component, substances with a pronounced trend, and random variables. This classification allows for better selection of forecasting and data transformation methods that can be more effectively applied to each class of substances.

The problem of environmental monitoring is formalized in two formulations: point and plane. The main stages of environmental monitoring are highlighted. These are collecting data on the state's history, monitoring the current state and predicting the state of environmental pollution in the future. Approaches and requirements for technical means at each stage are proposed. A review of known systems for monitoring air, water and soil pollution. The importance of the technical component is shown. Fundamental differences and new trends in using innovative technologies for monitoring environmental pollution parameters are identified.

A scientific hypothesis has been formulated that defines the author's vision of environmental monitoring organization in terms of combining software and hardware systems and using trend models to predict environmental pollution parameters. By formalizing the problem of environmental monitoring, the structure of the information system for environmental monitoring is proposed. The information system should include the following subsystems: a subsystem for collecting information about the state of the environment, a subsystem for storing and accumulating data, predicting the state of the environment, and a subsystem for user interaction.

It is indicated that constructing an air pollution monitoring system is also essential for the whole and safe operation of some critical infrastructure facilities, including power plants, processing and chemical plants, airports, tunnels and subways, etc. In case of poor environmental measurement near or inside these facilities, irreparable consequences for many people's environment, health and lives can occur.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Jinshui, S., Ming-Huang, W., & Yuh-Shan H. (2012). A historical review and bibliometric analysis of research on estuary pollution. Marine Pollution Bulletin, 64 (1), 13 – 21.

[2] Bellinger, Colin, Shazan, Mohomed, Jabbar, Mohomed, Zaiane, Osmar, & Osornio-Vargas, Alvaro. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. doi: 10.1186/s12889-017-4914-3. Retrieved from https://bmcpublichealth.biomedcentral.com/track/pdf/10.1186/s12889-017-4914-3.

[3] Government expenditure on environmental protection. (2019). Eurostat: Statistics Explained. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Government_expenditure_on_environmental_protection.

[4] Makridakis, S., Wheelwright, S.C., Hyndman, R.J. (2008). Forecasting: methods and applications. John wiley & sons, 656.

[5] Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., & et al. (2003). Extensive evaluation of neural network models for the prediction of no 2 and pm 10 concentrations, compared with a deterministic modelling system and measurements in central helsinki. Atmos Environ, 37(32), 4539–550. doi: 10.1016/S1352-2310(03)00583-1.

[6] Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J., Kolehmainen, M. (2004). Evolving the neural network model for forecasting air pollution time series. Eng Appl Artif Intell, 17(2), 159–67. doi: 10.1016/j.engappai.2004.02.002.

[7] Kolehmainen, M., Martikainen, H., Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. Atmos Environ, 35(5), 815–25. doi: 10.1016/S1352-2310(00)00385.

[8] Beckerman, BS., Jerrett, M., Martin, R.V., van Donkelaar, A., Ross, Z., Burnett, R.T. (2013). Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in california. Atmospheric Environ, 77, 172–177. doi: 10.1016/j.atmosenv.2013.04.024.

[9] Liu, B.C., Binaykia, A., Chang, P.C., Tiwari, M.K., Tsao, C.C. (2017). Urban air quality forecasting based on multi-dimensional collaborative support vector regression (svr): A case study of beijing-tianjin-shijiazhuang. PLOS ONE, 12(7), 1–17.

[10] Bobb, J.F., Valeri, L., Henn, B.C., Christiani, D.C., Wright, R.O., Mazumdar, M., & et al. (2014). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics, 16(3), 058.

[11] Gass, K., Klein, M., Chang, H.H., Flanders, W.D., Strickland, M.J. (2014). Classification and regression trees for epidemiologic research: an air pollution example. Environ Health, 13(1), 17. doi: 10.1186/1476-069X-13-17.
[12] Kuchansky, A., Biloshchytskyi, A., Andrashko, Y., Vatskel, V., Biloshchytska, S., Danchenko, O., Vatskel, I. (2018) Combined models for forecasting the air pollution level in infocommunication systems for the environment state monitoring. Proceedings of the 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems, IDAACS-SWS 2018, art. no. 8525608, pp. 125-130. DOI: 10.1109/IDAACS-SWS.2018.8525608.
[13] Kuchansky, A., Biloshchytskyi, A., Andrashko, Y., Biloshchytska, S., Honcharenko, T., Nikolenko, V. (2019) Fractal time series analysis in non-stationary environment. 2019 IEEE International Scientific-Practical Conference: Problems of Infocommunications Science and Technology, PIC S and T 2019 - Proceedings, art. no. 9061554, pp. 236-240. DOI: 10.1109/PICST47496.2019.9061554.
[14] Biloshchytskyi, A., Kuchansky, A., Biloshchytska, S., Dubnytska, A. (2017) Conceptual model of automatic system of near duplicates detection in electronic documents. 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2017 - Proceedings, art. no. 7916155, pp. 381-384. DOI: 10.1109/CADSM.2017.7916155.
[15] Biloshchytskyi, Andrii; Kuchansky, Alexander; Andrashko, Yurii; Neftissov, Alexander; Vatskel, Vladimir; Yedilkhan, Didar; Herych, Myroslava (2022) Building a model for choosing a strategy for reducing air pollution based on data predictive analysis. Eastern-European Journal of Enterprise Technologies, 3 (4-117), pp.23-30. DOI 10.15587/1729-4061.2022.259323.
[16] Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Biloshchytska, S., Danchenko, O. (2019) Development of Infocommunication System for Scientific Activity Administration of Educational Environment's Subjects 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, art. no. 8632036, pp. 369-372. DOI: 10.1109/INFOCOMMST.2018.8632036.
[17] Verbesselt J., Hydman R., Newnham G., Culvenor D. (2010). Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment. Volume 114, Issue 1, 15 January 2010, 106-115.
[18] Putrenko, V., Tihohod, V. (2015). Development of geostatistic models of intellectual analysis of air pollution. Eastern-European Journal of Enterprise Technologies, 4/10 (76), 21-26. doi: 10.15587/1729-4061.2015.47889.
[19] Clark, H. W. (2004). A Practical Geostatistics 2000. Publ. by Geostokos Ecosse, 440.
[20] Wackernagel, H. (2003). Multivariate Geostatistics. Springer, 403.
[21] Webster, R. (2000). Geostatistics for Environmental Scientists. John Wiley & Sons, 286.
[22] World Health Organisation. (2017). Data and Statistics; World Health Organisation: Geneva, Switzerland.
[23] Interactive map of monitoring environment of the environment. Retrieved from https://ecomapa.gov.ua.
[24] World's Air Pollution: Real-time Air Quality Index. Retrieved from http://waqi.info.
[25] Ahmed, A.N., Fazlul Haque, H.M., Rahman, A., Ashraf, M.S., Saha S., Shatabda S. (2017). A Participatory Sensing Framework for Environment Pollution Monitoring and Management. Computers and Society, 1701. 6429.
[26] Jong-Won Kwon, Yong-Man Park, Sang-Jun Koo, Hiesik Kim. (2007). Design of air pollution monitoring wystem using ZigBee networks for ubiquitous-city. 2007 International Conference on Convergence Information Technology, 1024-1031.
[27] Vasim, K., Ustad, A., Mali, S., Suhas S. Kibile. (2014). Zigbee based wireless air pollution monitoring system using low cost and energy efficient sensors. International Journal of Engineering Trends and Technology, 10/4, 456-460.

[28] Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Rickerby, D. (2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. Science of The Total Environment, 607, 691-705.
[29] Son, M., Cho, D. G., Lim, J. H., Park, J., Hong, S., Ko, H. J., Park, T. H. (2015). Real-time monitoring of geosmin and 2-methylisoborneol, representative odor compounds in water pollution using bioelectronic nose with human-like performance. Biosensors and Bioelectronics, 74, 199-206.
[30] Wernersson, A. S., Carere, M., Maggi, C., Tusil, P., Soldan, P., James, A., & et al. (2015). The European technical report on aquatic effect-based monitoring tools under the water framework directive. Environmental Sciences Europe, 27(1), 7.
[31] K. Kolesnikova, O. Mezentseva and T. Mukatayev, "Analysis of Bitcoin Transactions to Detect Illegal Transactions Using Convolutional Neural Networks," 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 2021, pp. 1-6, doi: 10.1109/SIST50301.2021.9465983.
[32] Olekh H., Kolesnikova K., Olekh T., Mezentseva O. Environmental impact assessment procedure as the implementation of the value approach in environmental projects (2021) CEUR Workshop Proceedings, 2851, pp. 206 – 216.
[33] K. Kolesnikova, T. Olekh, D. Lukianov and D. J. Obenewaa, "Markov Principles of Project Effectiveness," 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 2021, pp. 1-6, doi: 10.1109/SIST50301.2021.9465881.
[34] K. Kolesnikova, O. Mezentseva and O. Savielieva, "Modeling of Decision Making Strategies In Management of Steelmaking Processes," 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2019, pp. 455-460, doi: 10.1109/ATIT49449.2019.9030524.
[35] Chernova L., Titov S., Chernov S., Kolesnikova K., Chernova L., Gogunskii V. Development of a formal algorithm for the formulation of a dual linear optimization problem, (2019) Eastern-European Journal of Enterprise Technologies, 4 (4-100), pp. 28 – 36. DOI: 10.15587/1729-4061.2019.175105.
[36] Gogunskii V., Kolesnikov O., Oborska G., Moskaliuk A., Kolesnikova K., Harelik S., Lukianov D., Representation of project systems using the Markov chain, (2017) Eastern-European Journal of Enterprise Technologies, 2 (3-86), pp. 60 – 65. DOI: 10.15587/1729-4061.2017.97883.
[37] Honcharenko T., Akselrod R., Shpakov A., Khomenko O. (2023). Information system based on multi-value classification of fully connected neural network for construction management, IAES International Journal of Artificial Intelligence, 2023, № 12(2), pp. 593-601. DOI: 10.11591/ijai.v12.i2.pp593-601.
[38] Ryzhakova, G., Malykhina, O., Pokolenko, V., Rubtsova, O., Homenko, O., Nesterenko, I., Honcharenko, T. (2022). Construction Project Management with Digital Twin Information System. International Journal of Emerging Technology and Advanced Engineering, 12 (10), pp. 19-28. DOI: 10.46338/ijetae1022_03.
[39] Chernyshev, D., Ryzhakova, G., Honcharenko, T., Petrenko, H., Chupryna, I., Reznik, N. (2023). Digital Administration of the Project Based on the Concept of Smart Construction. Lecture Notes in Networks and Systems, 495 LNNS, pp. 1316-1331. DOI: 10.1007/978-3-031-08954-1_114.
[40] Chernyshev, D., Dolhopolov, S., Honcharenko, T., Haman, H., Ivanova, T., Zinchenko, M. (2022). Integration of Building Information Modeling and Artificial Intelligence Systems to Create a Digital Twin of the Construction Site. International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2022-November, pp. 36-39. DOI: 10.1109/CSIT56902.2022.10000717.
[41] Honcharenko, T., Mihaylenko, V., Borodavka, Y., Dolya, E., Savenko, V. (2021). Information tools for project management of the building territory at the stage of urban planning CEUR Workshop Proceedings, 2851, pp. 22-33.
[42] Dolhopolov, S., Honcharenko, T., Savenko, V., Balina, O., Bezklubenko, I., Liashchenko, T. (2023). Construction Site Modeling Objects Using Artificial Intelligence and BIM Technology: A

Multi-Stage Approach SIST 2023 - 2023 IEEE International Conference on Smart Information Systems and Technologies, Proceedings, pp. 174-179. DOI: 10.1109/SIST58284.2023.10223543.