DOI: 10.6084/m9.figshare.9788702
UDC 504:625.7

Yuanfang He
PhD student in the Department of Information Systems and Technologies, *orcid.org/0000-0002-6925-1540*
Taras Shevchenko National University of Kyiv, Kyiv

FOMALIZATION OF THE PROBLEM OF EVALUATION OF POLLUTION OF THE ENVIRONMENT

Abstract. *The paper considers the formalization of monitoring the state of the environment. The problem is considered in two statements: point and plane. The main stages have been highlighted: Collection of data on state history, monitoring of the current state and forecasting the state of environmental pollution in the future. Approaches and necessary requirements for technical means at each stage are offered. The analysis of the time series of air pollutants and the study of their trend, seasonal and random components justifies their classification of: substances with a pronounced seasonal component, substances with a pronounced trend and random variables. In accordance with this formalization, an information system for monitoring the state of the environment is proposed. The information system includes the following subsystems: collection of information on the state of the environment. data storage and accumulation, environmental forecasting and user interaction subsystem.*

Keywords: environment; pollution; information monitoring system

Introduction

Technical progress, the development of industry and agriculture every year increase the amount of waste. The main and most dangerous sources of pollution of the environment are human, given that the very person, the effects of its activities are radically affecting and changing the environment. Substances that pollute the environment can be solid (industrial dust), liquid (sewage) and gaseous (emissions of industrial gases) and have a harmful effect directly, after chemical transformations, or together with other substances.

The importance of the need for environmental research is also confirmed by the fact that governments in all leading countries spend an average of 0.8% of the budget of their countries (more than $ 600 billion) on environmental protection measures [1]. Among these, R & D spending is in 3rd place.

As stated in [2], each type of pollution (air pollution, pollution of underground and surface waters, soil contamination and impact on the biosphere) requires its own models and methods of research and forecasting. However, each of the authors of the studies studied has an understanding of the tasks of studying the state of the environment. That is why there is a need to formulate the task of monitoring the state of the environment from our point of view. Such formalization will make it possible to further investigate more ethically.

The purpose of the article

The research objective is:

1. To study approaches to the task of monitoring the state of the environment.

2. Formulate the task of monitoring the state of the environment.

Presenting main material

Setting the task of monitoring the state of the environment

Consider the task of evaluating environmental pollution in two formulas: point and plane.

So, in the article [3] for forecasting concentrations of NO2 and PM10 in urban conditions, the time series of the concentrations of the relevant substances were measured at two stations in the center of Helsinki. [4] addresses the use of weather stations to collect information on environmental pollution.

Suppose it is necessary to assess the pollution of the environment at a certain point. Environmental pollution can be estimated from a set of its indicators.

Let $R = (r_1, r_2, \dots, r_n)$, be a vector of real numbers that describes the state of the environment, where n is the number of indicators. Each coordinate of a vector is a certain indicator, for example, the concentration of sulfur dioxide or carbon monoxide in the air, the concentration of nitrates in water, etc. You can obtain relevant indicators using the appropriate technical means (weather stations, mobile and stationary sensors, etc.) and with the help of services.

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

That is $R(t) = (r_1(t), r_2(t), \dots, r_n(t))$, where *t* is a certain time. For the sake of simplification, we will assume that the indicators are updated over a certain period of time (hourly, daily, monthly). Then, without limiting generality, we will calculate time by a discrete value. That is $t_i = t_0 + \Delta t^{\circ} i$, where t_0 – the initial time from which the observation of the state of the environment is monitored; Δt – is the frequency of observation; $i = \overline{1,m}$, where m is the number of observations. Denote $r_j(t_i)$ through r_j^i. Then $R(t_i) = (r_1(t_i), r_2(t_i), \ldots, r_n(t_i)) = (r_1^i, r_2^i, \ldots, r_n^i)$.

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

1. Collect 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 problem it is necessary to build a certain database in which it is necessary to keep the history of the state of environmental pollution. There are two options for saving. The first way is to save the history of the environment as a set of dynamic rows, each of which reflects a change in one indicator. The second way is to preserve the sequence of the vector, each of which is a reflection of the state of the environment at a certain point in time.

To solve the second task, there is a need for a data source, a data transmission channel, and methods for converting information. The source of data on the state of the environment can be both hardware and so on other monitoring services of the environment. The data channel depends on the data source. The Internet is often the channel of transmission, but sometimes there is a need for data transmission through service protocols, such as Zigbee [5; 6], to a kind that can be stored in the system described in the first task.

Consider the third problem for the case when it is necessary to predict only one indicator. Then the prediction problem is to calculate the values of the pollutant index with the horizon θ> 1, that is, for each time moment m+1, m+2, …m+ θ. In other words, it is necessary to prolong the dynamic series of pollution indicators. $R^* = (\overline{r}_{n+1}, \overline{r}_{n+2}, \ldots, \overline{r}_{n+\theta})$, where the horizon θ is fixed to the calculation of the forecast.

Let p be the size of the retrospective sample, that is, the size of the region of the dynamic series, followed by the point at which the forecast (point t_m) is calculated, and which participates in the calculation of the predictive values for p <m. The functional dependence, on the basis of which the values are predicted, is called the forecasting model.

Moreover, $\overline{r}_{n+\tau}$ is the predicted score calculated at r_n for τ points ahead with the period $\tau = \overline{1,\theta}$. If to formally designate a model as f, then the forecast calculated at r_n for one forward point or with period 1 can be defined as follows: $\overline{r}_{n+1} = f(r_{n-m+1}, r_{n-m}, \ldots, r_n)$.

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

In the articles [7,8], the time series of 16 air pollutants were analyzed and their trend, season and random components were analyzed (Table 1).

*Table 1 – **Air pollutants***

№	Pollutant	Components		
		Trend	Seasonal	Random
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	Nitric 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	Ethylbenzene	0,0907	0,3819	0,5274
14	Trichloromethane	0,0659	0,1296	0,8045
15	Benzopyrene	0,0237	0,4878	0,4886
16	Tetrachloromethane	0,0194	0,4352	0,5454

According to the revealed regularities in the dynamic rows of pollutants can be divided into 3 classes:

1. Substances with pronounced seasonal component: benzopyrene, sulfur dioxide, carbon monoxide. The presence of the cycle is due to the fact that in the winter period, the emissions of these pollutants from heat power plants, motor vehicles considerably increase. Summer and winter periods affect the concentration of airborne contaminants.
2. Substances with a pronounced trend: benzene, toluene, ethylbenzene, nitric oxide. In these pollutants, in addition to the seasonal component of Jaskar, a trend is expressed to increase concentration.
3. Random values in which it is difficult to allocate the seasonal component: trichloromethane, ammonia. The impact on their level is provided by accidental events (non-periodic processes, volley and accidental emissions, adverse meteorological conditions, etc.).

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

Taking into account the above, we can assume that the use of prediction of the level of air pollution on models based on neural networks is a good option. Neural networks are able to take into account hidden dependencies. The basis of the formation of samples for training and testing of neural networks are dynamic series.

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

The key difference between this statement is the presence of an entire network of surveillance. Then, information on the state of the environment can be described as a set of tuples $\langle R_i, C_1\rangle$, where R_i is a vector that displays the state of the environmental pollution indicators at time t_i,, and C_1 is the location information where the relevant data is received. Set in a certain coordinate system.

The application of Land use regression (LUR) is described in [9]. LUR is essentially an interpolation technique that uses the polluter's interest as a dependent variable, using landscape data, traffic, and physical environmental parameters that are used as independent predictors.

The research [10] considers contaminated with urban and rural pollutants in the Nansha River. This study first took into account the contamination of substances from point sources and non-point sources in the watershed of the river.

In the article [11], Urban Air Quality Index (AQI) forecasting in China uses Support Vector Regression (SVR). It is shown that geographic location plays a significant role in the AQI forecast.

It should also be taken into account that much of the methods of forecasting and finding the existence of dependencies between magnitudes on the plane are constructed from the calculation of the coordinates given in the Cartesian system. Observation of the state of the environment has an effect on geographic coordinates.The system of geographic coordinates is used to determine the position of points of the earth's surface relative to the equator and the initial (zero) meridian. The coordinates are the angular dimensions: the geographical latitude B and the geographical longitude L. Longitude (the angle between the meridian plane at the observation point and the zero (Greenwich) meridian), Latitude (the angle between the slope line and the equatorial plane) determine the position of the point on the Earth's surface. Measured in degrees (°), longitude - from 0 ° to 180 ° west and east of Greenwich, latitude from 0 ° to 90 ° north, from 0 ° to -90 ° south of the equator.

The system of geographic coordinates is a spherical coordinate system. Therefore, to apply the transition formula to the Cartesian system.

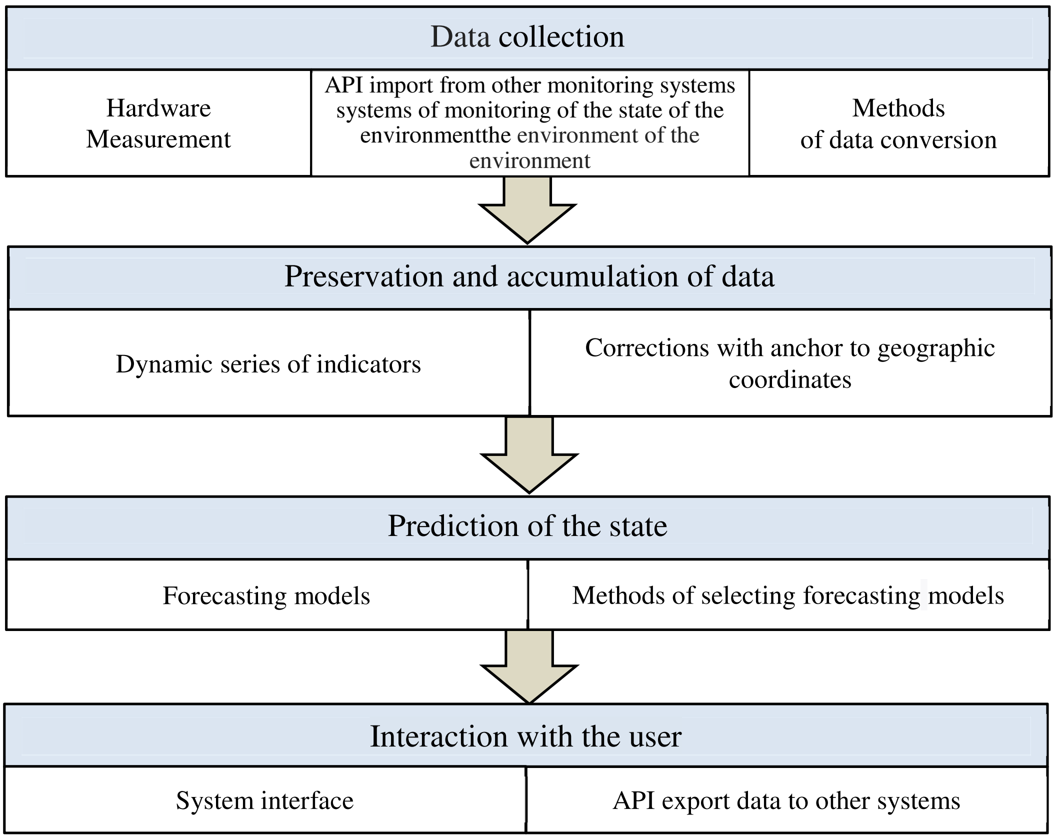

Figure 1 – Conceptual diagram of the monitoring system of the environment

Information system for monitoring of the environment

Taking into account the above mentioned information system for monitoring the state of the environment should include such subsystems (fig. 1):

1. Subsystems for collecting information on the state of the environment. This subsystem includes Hardware Measurement of Environmental Statements, Import APIs from other environmental monitoring systems and methods for converting data into a single format that is used in the data storage subsystem.
2. The subsystem of data storage and storage should be optimized taking into account the specifics of the data stored.
3. The subsystem of forecasting of the environment includes prediction models and methods of choosing which of the models should be used in a particular case to achieve greater prediction accuracy.
4. Subsystem interaction with the user is one of the most important parts of the information system. It must provide information in a convenient form. In particular, the submission of reports, interactive maps of the state of the environment, recommendations on the dangerous changes of environmental factors, such as exceeding the maximum permissible concentrations of certain pollutants.

Each subsystem can be considered as a separate module. The modular structure of the system will allow to expand and modify the capabilities of each of the modules independently of others. Also, the modular structure increases the stability and flexibility of the system. And given the modern approach to software development, the modular approach allows us to implement a micro-service approach when a system consists of a set of independent microservices.

Conclusions and perspectives of further research

The article formalizes the task of monitoring the state of the environment in two formulas: point and plane. The main stages of monitoring the state of the environment are highlighted.

The analysis of the time series of air pollutants and the study of their trend, seasonal and random components justifies the need for classifying and developing forecasting methods and models for each class.

Consequently, it can be concluded that the formalization of the task of monitoring the state of the environment generates a unified understanding of the problem, which will allow more ethical further research.

References

1. 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

2. Yuanfang, He & Vatskel, Igor. (2019). Problem of evaluation of pollution of the environment. Management of development of complex systems, 37, 168–172.

3. 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

4. Kuchansky, A., Biloshchytskyi, A., Andrashko, Yu., Vatskel, V., Biloshchytska, S., Danchenko, O. & et al. (2018). Combined models for forecasting the air pollution level in infocommunication systems for the environment state monitoring. 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS). Lviv, 125–130. DOI: 10.1109/IDAACS-SWS.2018.8525608

5. Kwon,Jong-Won, Park,Yong-Man, Koo, Sang-Jun & Kim, Hiesik. (2007). Design of air pollution monitoring wystem using ZigBee networks for ubiquitous-city. 2007 International Conference on Convergence Information Technology, 1024-1031.

6. Vasim, K., Ustad, A., Mali, S. & Kibile, Suhas S. (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.

7. Panchenko, A.A. (2018). Forecating the level of air pollution of the city with the development of chemical and petrochemical industry (on the example of sterlitamak of the republic of Bashkortostan). Extended abstract of candidate's thesis. Ufa.

8. 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

9. 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, 77:172–7. doi: 10.1016/j.atmosenv.2013.04.024

10. Jia, H., Wang, S., Wei, M. & Zhang, Y. (2011). Scenario analysis of water pollution control in the typical peri-urban river using a coupled hydrodynamic-water quality model. Frontiers of Environmental Science & Engineering in China, 5(2), 255–265.

11. 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. [PMC free article] [PubMed]

Стаття надійшла до редколегії 07.03.2019

Юфнгфанг Гі
Аспірант кафедри інформаційних систем і технологій, *orcid.org/0000-0002-6925-1540*
Київський національний університет імені Тараса Шевченка, Київ

ФОРМАЛІЗАЦІЯ ЗАДАЧІ ОЦІНЮВАННЯ ЗАБРУДНЕНОСТІ НАВКОЛИШНЬОГО СЕРЕДОВИЩА

Анотація. *Розглянуто формалізацію моніторингу стану навколишнього середовища. Задача розглядається в двох постановках: точковій і на площині. Виділено основні етапи: збирання даних про історію стану спостереження за поточним станом та прогнозування стану забруднення навколишнього середовища в майбутньому. Запропоновано підходи та необхідні вимоги щодо технічних засобів на кожному із етапів. Аналіз часових рядів забрудників повітря та дослідження їх трендової, сезонної і випадкової складових обґрунтовує класифікацію їх на: речовини з ярко вираженою сезонною складовою, речовини із вираженим трендом та випадкові величини. Відповідно до цієї формалізації запропонована структура – це інформаційна система моніторингу стану навколишнього середовища. Інформаційна система включає такі підсистеми: збирання інформації про стан навколишнього середовища, збереження та накопичення даних, прогнозування стану навколишнього середовища та підсистема взаємодії із користувачем.*

Ключові слова: навколишнє середовище; забруднення; інформаційна система моніторингу

Link to publication

APA *Yuanfang, He, (2019). Fomalization of the problem of evaluation of pollution of the environment. Management of development of complex systems, 38, 168 – 172, dx.doi.org\10.6084/m9.figshare.9788702.*

ДСТУ *Юфнгфанг Гі. Формалізація задачі оцінювання забрудненості навколишнього середовища [Текст] / Юфнгфанг Гі // Управління розвитком складних систем. – № 38. – 2019. – С. 168 –172, dx.doi.org\10.6084/m9.figshare.9788702.*