

ABSTRACT

of the doctoral thesis for the degree of Doctor of Philosophy (PhD) in the specialty:
6D070300 - «Information systems (by industry)»

BUGUBAYEVA ALINA ZHANATBEKOVNA

DATA MINING METHODS AND ALGORITHMS FOR AIR POLLUTION MONITORING SYSTEM

General description of work. The dissertation work is devoted to the development of mathematical and intellectual support for an information system that implements the processes of forecasting and decision support in the field of atmospheric air monitoring, based on the use of methods and algorithms for data mining. The proposed methods make it possible to assess and predict the negative impact of industrial emissions on public health. Based on the developed methods and algorithms for data mining, the architecture of an intelligent information system and its software implementation were designed.

Keywords: atmospheric pollution monitoring, data mining technology, simulation modeling, neural network modeling, intelligent information system architecture.

The relevance of research. The prosperity and competitiveness of the Republic of Kazakhstan against the world trend, technological innovations and scientific achievements is associated with the high-quality development, implementation and effective implementation of state programs, strategic plans and projects covering all spheres of human activity. For example, the national project "Green Kazakhstan" is aimed at creating a favorable living environment for the population and improving the environmental situation, including: improving the quality of atmospheric air, efficient handling of production and consumption waste, efficient and careful use of water, modernization of the ecological consciousness of the population.

Environmental processes are complex and non-linear, which complicates the processes of forecasting and assessing environmental performance. In the context of the rapid development of information technology, along with the classical mathematical apparatus for modeling environmental processes, the use of effective methods of data mining allows generating and accumulating new knowledge by identifying hidden patterns.

The most important criteria for the quality of atmospheric air monitoring are accuracy and reliability, and in order to improve the quality, it is necessary to use intelligent algorithms to reduce the risks associated with measurement errors. An important block in research related to environmental issues, including air pollution, is the issue of the impact of pollution on human health, but since this is a complex multi-parameter process, it is important to build an integrated model that takes into account the most important factors of influence.

Thus, in order to form a set of scientifically based practical recommendations that will be used to make optimal decisions in the regional environmental management system, effective methods, algorithms and tools for intellectual support are needed on the platform of intensively developing tools in this area.

The main idea of this study. The dissertation research is aimed at obtaining the most complete picture of monitoring the state of the object, which requires the development of models, methods and numerical methods, as well as state algorithms based on the measurement of significant parameters of the object, including incomplete, imperfect and sloppy data, for their use in decision support. These models can be implemented using artificial intelligence systems, fuzzy logic apparatus, neural networks, genetic algorithms, and combinations thereof. The neural network model can be instructed by measuring the data and allows the summing of incomplete noisy data. Fuzzy logic methods and neural networks are able to process high quality information, which is their advantage.

The object of the study is the process of monitoring the level of air pollution in the urban environment.

The subject of the study is the methods and algorithms for data mining of urban air pollution.

Purpose of the work: development of methods and algorithms for data mining for the atmospheric pollution monitoring system in order to increase the reliability of multi-parameter environmental control.

To achieve this goal, the following **main tasks** are formulated in the work:

1. Development of software for the control and decision-making system in the process of environmental monitoring based on the construction of a multifactorial model.
2. Development of a simulation model for quantitative assessment of the quality of instrumental control in the city air monitoring system;
3. Development of an algorithm for intelligent processing of environmental data for the control and decision-making process.
4. Development of a neural network model for data analysis for the decision-making process in a multi-parameter environment.
5. Development of the architecture of an intelligent information system for assessing the negative impact of industrial emissions on public health.

Basic research methods. The methodological basis of the study is a systematic approach. In the work, to solve the tasks set, the apparatus of the theory of systems management, methods of statistical analysis, artificial intelligence, and information system design technologies are used.

The scientific novelty of the dissertation research lies in the fact that for the first time a methodology for data mining in the atmospheric pollution monitoring system was proposed, including the development of mathematical and intellectual support for an information system that implements the processes of assessing and predicting the impact of atmospheric pollution on public health.

Scientific provisions submitted for defense:

1. Methodology for assessing and predicting the quality of public health based on the construction of an integrated multifactorial model.
2. Simulation model and algorithm for quantitative assessment of the quality of instrumental control in the city air monitoring system.
3. Architecture of an intelligent information system for assessing the negative impact of industrial emissions on public health based on the use of neural network technology.

The scientific and practical significance of the work is provided by the developed software product "Intellectual information system for assessing the negative impact of industrial emissions on human health (on the example of Ust-Kamenogorsk data)" with certifying intellectual property documents.

The results of the study can be used in the monitoring systems of industrial cities, subject to the negative impact of emissions from industrial enterprises, to improve the efficiency and reliability of atmospheric air quality monitoring.

Approbation of work. The main provisions and results of the work were reported and approved at the following international and scientific conferences:

1. III International scientific and practical conference "Global science and innovations 2018: Central Asia", Nur-Sultan, Kazakhstan.
2. LXVIII International correspondence scientific and practical conference "European research: innovation in science, education and technology" London, England.
3. "Communications in Computer and Information Science", Ust-Kamenogorsk, Kazakhstan.
4. "Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments", Wilga, Poland.
5. "The International Society for Optical Engineering", Wilga, Poland.

Publications. The results obtained in the dissertation are published in 11 papers, including 2 articles in journals peer-reviewed in the Scopus database, such as: Journal of Theoretical and Applied Information Technology, Pakistan (CiteScore percentile 34%); Przegląd Elektrotechniczny (CiteScore percentile 22%), 3 articles in the publication recommended by the Committee for Quality Assurance in Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan, Bulletin of the East Kazakhstan Technical University. D. Serikbaeva, 6 works in collections of materials of international conferences (4 of which are reviewed in the Scopus database). There is also 1 certificate of state registration for the object of copyright (computer program) Intelligent information system for assessing the negative impact of industrial emissions on public health (using the data of the city of Ust-Kamenogorsk as an example) No. 16777 dated 04/20/2021.

The structure and scope of the dissertation. The dissertation consists of an introduction, 4 chapters, a conclusion, a list of references from 106 titles presented on 89 pages of computer text, includes 43 figures, 5 tables and 2 appendices.

The introduction substantiates the relevance of the research topic, the purpose, object, subject, tasks and methods of research, scientific novelty, scientific provisions, practical value and implementation of the results of the work, provides information about publications and approbation of the work.

The first section of the thesis contains an analytical study in the form of a review of theoretical and practical approaches to solving problems on the subject under study. It is indicated that one of the promising and poorly studied areas of application of artificial intelligence is ecology, and at present ecology is acquiring the role of an integrator of health problems and human economic activity.

The analysis of digital transformation tools in the field of environmental issues and the development of the hardware and software component of environmental monitoring systems are carried out, showing that methods and means of intelligent data processing are currently relevant.

The section provides an overview of the existing mathematical software for air pollution monitoring systems, showing that most studies use a mathematical apparatus based on differential equations and methods of mathematical statistics, and also substantiates the need for effective methods that allow generating and accumulating new knowledge by identifying hidden patterns.

In the second section, an analysis of the objective prerequisites for the system intellectualization of formal decision-making processes in a multi-parameter environment was carried out, it was revealed that the system being developed must necessarily contain a block for quantitative risk assessment in the form of monitoring.

As a result of the above analysis, the main limitations and shortcomings of the above methods were identified and an approach was proposed for implementing the process of intellectual information processing.

As a result of a statistical experiment, significant factors influencing life expectancy and quality of health were identified. Using the methodology of expert assessments and the automation program for this technology, the weights of each of the listed health factors were found.

A methodology for assessing and predicting the quality of public health has been developed, including mathematical support for the control and decision-making system in the process of environmental monitoring based on the construction of an integrated multifactorial model.

In the third section, a simulation model and an algorithm for quantifying the quality of instrumental control in the city air monitoring system are developed. A neural network model for data analysis has been developed to support decision making in a multi-parameter environment. A method for intelligent processing of atmospheric air pollution measurement data has been developed, including a simulation model for assessing the risks of measurement errors and a neural network model for assessing the impact of air pollution on public health.

The fourth section describes the architecture of an intelligent information system for assessing the negative impact of industrial emissions on human health. The structure

and operation algorithm of the developed software of the intelligent information system are described.

At the end of the dissertation work, a list of the main results and conclusions of the dissertation research is given, on the basis of which the provisions submitted for defense are confirmed, an assessment of the scientific novelty and practical significance of the research is given.

The content of the dissertation ends with a list of references and applications.