Ministry of Science and Higher Education of the Republic of Kazakhstan

SDU University

Bekzat Makhanbet

Automatic Monitoring System of emissions

THESIS

Presented in Partial Fulfilment for the

Degree of Master of Technical Science in Computer Science

(degree code: 7M06102)

Department of Computer Science

Faculty of Engineering and Natural Sciences

Supervisor: **Dauren Ayazbayev**

Kaskelen, June 2024

SDU University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty of Engineering and Natural Sciences

Assistant Professor, PhD Akhmedov Ramis

____

” _____” ______________________________ 2024

Topic of the thesis:

Automatic Monitoring System of emissions

Thesis submitted as part of the requirements for the award of the MSc in “7M06102 - Computer Science”, SDU University

Head of Department Zhanar Mukash

Academic Supervisor Dauren Ayazbayev

Master student Bekzat Makhanbet

Kaskelen, 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Bekzat Makhanbet

June 2024

Acknowledgements

My curator I would like to thank my supervisors prof. Dauren Ayazbayev for his support and assistance in my work. Without him part of this research work would be impossible

My colleagues - ecologists I also want to thank my colleagues, honored ecologists Dairova Meruyert and Marina Kim for chewing on aspects of ecology for me so that I could digitize this area. I work with these charming people. Together we did lots of work for area of Green tech in Kazakhstan.

Colleagues from Russia The big part of help in area of programming ecological business processes and simulating model of OND-86 I got from our colleagues from Russia company "Integral". They are on of biggest green tech companies in Russia and captured market by 80 percent. They helped me since we are very close partners with them.

Dedication

This thesis is dedicated to:

Ministry of Ecology of Kazakhstan, to users of Kazakhstan natural resources and to all ecologists

Abstract

"Sergeks" in area of Ecology

Our country is very polluted because of the incorrect ecological laws, old manufactures and irresponsible users of natural resources. And our the state has no tools to control it. Our research work will solve this problem by using informational technologies. It is very important to constantly evaluate the air quality in urban areas to warn residents to risks posed by air they breathe. However, the construction typical monitoring networks in poor countries will impractical due to their enormous cost. Nationwide Air Quality Network monitoring was created to offer air quality indicators, and this is important for citizen to be informed about the present and upcoming air quality in its current place to avoid overexposure. This network made it possible to broadcast frequently quality control and development the most important information system that helps prevent dangerous circumstances. Difficulty in collection and interpretation of time series and Direct data is fundamental disadvantage of an unusual network topology.

Old mathematical simulation model

Actually in this area I acted more like as a developer. We had collabration of 2 ecological companies of Kazakhstan and Russia, "Koktem Tech" where I am tech lead of company and "Integral" Russian biggest green tech company. Little bit about current situation of environmental permits of Kazakhstan, we are currently using model approved in 1986 named OND-86. This model does not coincide with current realities and gives incorrect sanitary protection zones for residential areas, which could lead to an environmental disaster. Our country uses at this moment computer Russian program called "ERA" which they themselves have not used since 2013. But for out Ministry of ecology doesnt see ant problem. We will also try to solve this problem and develop new desktop or web application which will correspond to current realities.

Аңдатпа

Экология саласындағы "Сергек"

Біздің еліміз экологиялық заңдардың дұрыс еместігінен, ескі өндірістер мен табиғи ресурстарды жауапсыз пайдаланушылардың кесірінен қатты ластанған. Ал біздің мемлекеттің оны бақылайтын құралы жоқ. Біздің зерттеу жұмысымыз бұл мәселені ақпараттық технологияларды қолдану арқылы шешеді. Тұрғындарды тыныс алатын ауадан туындайтын қауіптер туралы ескерту үшін қалалық жерлерде ауаның сапасын үнемі бағалау өте маңызды. Дегенмен, кедей елдерде типтік бақылау желілерін салу олардың орасан зор құнына байланысты мүмкін болмайды. Ауа сапасының жалпы ұлттық желісінің мониторингі ауа сапасының көрсеткіштерін ұсыну үшін құрылды және бұл азаматтарға шамадан тыс әсер етпеу үшін қазіргі және алдағы ауа сапасы туралы хабардар болуы маңызды. Бұл желі сапаны бақылауды жиі таратуға және қауіпті жағдайлардың алдын алуға көмектесетін ең маңызды ақпараттық жүйені әзірлеуге мүмкіндік берді. Уақыттық қатарларды және тікелей деректерді жинау және интерпретациялау қиындығы әдеттен тыс желі топологиясының негізгі кемшілігі болып табылады.

Ескі математикалық модельдеу моделі

Іс жүзінде бұл салада мен әзірлеуші ретінде әрекет еттім. Бізде Қазақстан мен Ресейдің 2 экологиялық компаниясы, «Көктем Тех» және мен компанияның техникалық жетекшісімін және «Интеграл» ресейлік ең ірі жасыл технология компаниясымен ынтымақтастығымыз бар. Қазақстанның экологиялық рұқсаттарының қазіргі жағдайы туралы аздап айтсақ, біз қазір 1986 жылы бекітілген ОНД-86 деп аталатын үлгіні қолданамыз. Бұл модель қазіргі шындыққа сәйкес келмейді және экологиялық апатқа әкелуі мүмкін тұрғын аудандар үшін дұрыс емес санитарлық қорғау аймақтарын береді. Біздің ел қазір 2013 жылдан бері пайдаланбай келе жатқан «ЭРА» атты ресейлік компьютерлік бағдарламаны пайдалануда. Бірақ экология министрлігінде құмырсқа мәселесі байқалмайды. Біз сондай-ақ осы мәселені шешуге тырысамыз және ағымдағы шындыққа сәйкес келетін жаңа жұмыс үстелі немесе веб-қосымшаны әзірлеуге тырысамыз.

Аннотация

«Сергеки» в области Экологии

Наша страна очень загрязнена из-за неправильных экологических законов, старых производств и безответственных пользователей природных ресурсов. И у нашего государства нет инструментов, чтобы это контролировать. Наша исследовательская работа позволит решить эту проблему с помощью информационных технологий. Очень важно постоянно оценивать качество воздуха в городских районах, чтобы предупреждать жителей о рисках, связанных с воздухом, которым они дышат. Однако строительство типичных сетей мониторинга в бедных странах будет нецелесообразным из-за их огромной стоимости. Национальная сеть мониторинга качества воздуха была создана для предоставления показателей качества воздуха, и это важно, чтобы граждане были информированы о текущем и предстоящем качестве воздуха в его текущем месте, чтобы избежать чрезмерного воздействия. Эта сеть позволила часто транслировать контроль качества и разработку важнейшей информационной системы, помогающей предотвращать опасные обстоятельства. Сложность сбора и интерпретации временных рядов и прямых данных является фундаментальным недостатком необычной топологии сети.

Старая математическая модель моделирования

Собственно в этой области я выступал скорее как разработчик. У нас было сотрудничество двух экологических компаний Казахстана и России: «Коктем Тех», где я являюсь техническим руководителем компании, и «Интеграл», крупнейшей российской компании в сфере экологических технологий. Немного о текущей ситуации с экологическими разрешениями в Казахстане, сейчас мы используем модель ОНД-86, утвержденную в 1986 году. Эта модель не соответствует современным реалиям и дает неверные санитарно-защитные зоны жилых территорий, что может привести к экологической катастрофе. В нашей стране сейчас используется компьютерная российская программа под названием «ЭРА», которую они сами не используют с 2013 года. Но наше Минэкологии не видит проблемы с муравьями. Мы также постараемся решить эту проблему и разработать новое настольное или веб-приложение, которое будет соответствовать текущим реалиям.

Abbreviations

OND-86 - General Regulatory Documentation 1986
AMS - Automatical monitoring system
VOCs - Volatie organic compounds
LOOC - Leave-one-out crossvalidation
ERA - Environmental calculations association
JSC - Joint-stock company
PES - Permissible emission standards
MEGPR - Ministry of Ecology, Geology and Natural Resources
EES - Environmental ecological system
EEP - Environmental ecological portal

Table of Contents

Chapter 1

Background and motivations

1.1 Introduction

An air pollutant is a substance in the atmosphere that has the potential to be harmful to both humans and the environment. The substance might be vapors, droplets of liquid, or solid particles. A pollutant may be produced by either nature or by people. There are primary and minor categories for pollutants. A process, like the eruption of a volcano, often produces primary pollutants. Other examples are carbon monoxide gas from automotive exhaust or sulfur dioxide released by industry. Secondary pollutants are not directly released. Instead, they arise in the air as a consequence of fundamental contaminants interacting with one another. Ozone at ground level is a well-known illustration of a secondary contaminant. Pollutants are created, either directly or via the production of other major pollutants. Methane is a potent greenhouse gas that hastens the process of global warming. Other hydrocarbon VOCs are also significant greenhouse gases because they prolong the atmospheric lifespan of methane and contribute to ozone production. This influence varies according on the quality of the nearby air. The quantity of these pollutants in the air has an effect on the happiness and health of city dwellers. When pollution levels are high, it is advisable to reduce or cease participating in outdoor activities and exercises. Particular attention should be paid to this while dealing with young children, the elderly, and those who have respiratory or cardiac issues. In this regard, the general public should have quick and easy access to relevant information on the degree of air pollution in metropolitan areas.

For effective environmental and health protection, a reliable air quality monitoring system is a need. The system must be simple to use, trustworthy, delicate, and affordable. The device also has to be very sensitive to low concentrations of gaseous air contaminants like carbon monoxide and hydrogen, both of which are found in cigarette smoke. The current trend for the development of air contaminants monitoring and alarm systems is to increase sensitivity and minimize response time, particularly at low air pollutants concentrations. For these compounds, the conventional environmental sampling methods include hand-collected grab samples that are subsequently removed off-site and forwarded to a lab for analysis.A lot of things, including an gas sensors made of semiconducting materials are now of great interest to both sensor users and scientists. Tin dioxide, a prototype material, and metal oxide gas sensors in general have been the subject of several publications that have been published in the literature.

The study of the state of atmospheric air in the cities of Kazakhstan as a whole is carried out mainly in large cities, however, an analysis of the current situation shows that monitoring of the air environment of small cities and remote industrial regions is not carried out in full. Limited capabilities for environmental control, incompetence of specialists, irresponsible attitude towards the environment of the majority of industrial enterprises in the region with a post-Soviet heritage, as an example of which in the current days, we can observe in the case of Arcelor Mittal JSC, as ensuing circumstances associated with this state - unreliability of data on the inventory of pollutant emission sources and the difficulties of collecting scattered and incomplete information located in organizations of various departmental affiliations, as well as the lack of analytical information centers with the necessary equipment (not all state laboratories of territorial environmental departments are fully equipped, depending on the need for this or that equipment) and software (which in general is a big problem not only of its absence, but also of the technical capabilities of its application) necessitate the creation of a universal geographic information system for a qualitative assessment of air pollution in the industrial regions of Kazakhstan in the first place.

1.2 State planning in the field of use of natural resources

The subject of the study is the process of formation of strategic planning tools in the field of environmental management in the Republic of Kazakhstan. The purpose of the study is to identify organizational problems in the formation of an institute of state strategic planning in the field of use and protection of natural resources, as well as to formulate scientifically based proposals for overcoming them using IT technologies.

Environmental management planning is an element of rational management in the field of environmental management and environmental protection. The purpose of planning is to use the country's natural resource potential comprehensively and economically. The goals of planning in market conditions are preserved, but the search for new forms and methods of planning continues, its tasks and functions change.

In the field of environmental protection and rational use of natural resources, the strategic goals of the Republic of Kazakhstan are as follows: 1) solving problems of development of the country's economic complex, taking into account the environmental and natural-geographical conditions of each specific territory; 2) achieving environmental quality in each specific territory; 3) preservation and restoration of the genetic fund of the animal world and biosphere balance; 4) using the entire natural resource potential of the country rationally.

To achieve these strategic goals, environmental management planning is necessary, which represents a whole set of actions and decisions, and the development of specific strategies.

Planning begins with the collection and analysis of information in order to predict future changes in the situation, the development of specific activities, the formation of an organizational structure, the allocation of resources, and the selection of sources of financing.

Programs, plans and activities for environmental protection and rational use of natural resources are specifically developed in the field of environmentally safe development of production and deployment of productive forces, in the field of creating a healthy living environment for the population, in the field of preventing and reducing man-made accidents and disasters.

An important point is the formation of a reliable database on the state of the country's natural resource potential. In order for executive authorities and local governments to provide reliable information about the state of natural resource potential in the country, a system of comprehensive territorial cadastres of natural resources is being formed. To accomplish this task, it is necessary to collect and process a lot of data, systematize it and regularly update it. For these purposes, it is necessary to develop existing IT technologies.

1.3 Aims of research

Goals of creating the System:
ensuring the reception of data from automated systems for monitoring emissions into the environment of natural resource users;
submission and acceptance of industrial environmental control reports; definition of object categories;
collection, accumulation and systematization of the state fund of environmental information;
monitoring releases and transfers of pollutants;
updating indicators and modifying data collection forms, algorithms for processing and presenting data in accordance with changes in norms and requirements of environmental legislation.

The creation of any information system, like a database, is always a rather complex process that consists of many successive stages. It is necessary to highlight the main technical tasks and the sequence of their solutions:
Task 1. Determination of a priority list of pollutants, the concentration of which should be assessed online for the region and a list of effective sources of pollution.
Task 2. Preparation of a system for transmitting, receiving and loading operational data from meteorological observations and industrial environmental control data
Task 3. Preparation of data on sources of air pollution in the region.
Task 4. Creating a database

During operation, the information system must perform the following types of tasks:
- monitoring the impact on the natural environment and supporting the adoption of management, design, technical, technological and investment decisions in the field of environmental activities;

- formation of a database of laboratory measurements at the border of the sanitary protection zone of large enterprises and at sources of emissions and sampling at wastewater outlets, at a control point, in an observation well, before and after treatment facilities with the possibility of generating internal and external reporting documentation, and comparative analysis of data for a long period;
- automated calculation of emissions and emissions of greenhouse gases into the environment in the context of an individual enterprise and group of companies (consolidated volumes of PES);
- optimization of the process of generating environmental reporting for enterprises;
- formation of a database of environmental indicators of production facilities/sites, including indicators of the use and withdrawal of natural resources;
- visualization of the dynamics of environmental impact through analytical attributes (graphs, diagrams, tables, etc.);
- monitoring the current state of the environment and natural resources, monitoring pollution of atmospheric air, water resources and land (surface and subsoil) by emissions, discharges and waste from industrial enterprises;
- providing the required environmental information to state authorities and local governments, citizens, enterprises, institutions and organizations.

The main tasks when creating a database are:
Ensuring that all necessary information is stored in the database.
Ensuring the ability to obtain data for all necessary requests.
Reduce data redundancy and duplication.
Ensuring data integrity (correctness of their content): eliminating contradictions in data content, eliminating their loss, etc.
Ensuring system functionality
Ensuring system scalability
Providing the ability to migrate information and integrate it
When identifying the subject area at the design stage, a conceptual database model was created, on which objects and connections between them were represented.
Next, a logical model was created defining the data structure and relationships between them. Data sources are enterprises in the region, which fill out the presented form on a digital platform and send the data to the database. The entire information base is stored on a cloud server. The database is managed through a DBMS. Users usually access the database not directly through the DBMS, but using an external interface - an application included in the AIS. Analytical information upon user request is displayed on the external interface.

During the collection of data from various sources, the main processes of the information system were identified, on the basis of which the corresponding modules required for the operation of the system were created. The modular design principle allows for optimal system operation without additional loads. The developed database system must consists of 6 main modules: data input, meteorological parameters, calculation module, data analysis, visualization, data output. Each module solves specific problems for monitoring and predictive analysis of atmospheric air pollution.

Chapter 2

MATERIALS AND METHODS OF RESEARCH

2.1 FUNDAMENTALS OF ENVIRONMENTAL REGULATION

According to the Environmental Code of the Republic of Kazakhstan, the instruments of state regulation in the field of environmental protection are:

1) licensing of activities in the field of environmental protection; environmental regulation;

2) technical regulation in the field of environmental protection; environmental assessment;

3) state environmental assessment;

environmental permits and environmental impact declarations;

4) monitoring of the environment and natural resources;

5) state environmental control;

6) notification procedure for collecting, sorting and (or) transporting waste;

7) instruments of state regulation in the field of emissions and absorption of greenhouse gases.

At the moment, economic and other activities for which an environmental impact assessment is carried out are divided into 4 categories according to the significance and completeness of the assessment - I, II, III, IV.

Category I includes objects that have a significant negative impact on the environment.

Category II includes facilities that have a moderate negative impact on the environment.

Category III includes facilities that have an insignificant negative impact on the environment.

Category IV includes facilities that have minimal negative impact on the environment.

Activities that do not belong to hazard classes according to the Environmental Code or do not meet the criteria set out therein are classified as category IV objects.

Our research work mostly will be directed to digitize harm for our ecology from companies which will be in the list of I and II category.

2.2 State environmental control

State environmental control is the activity of the authorized body in the field of environmental protection, aimed at ensuring compliance by individuals and legal entities with the requirements of the environmental legislation of the Republic of Kazakhstan. State environmental control is carried out in the following areas:

1) compliance with the provisions of the Environmental Code in the field of environmental protection;

2) compliance with environmental requirements in the field of specially protected natural areas;

3) compliance with environmental requirements during conservation and liquidation of the consequences of subsoil use operations, reclamation of disturbed lands;

4) fulfillment of extended obligations of manufacturers (importers);

5) fulfillment by the operator of the extended obligations of manufacturers (importers) of the requirements determined by this Code;

6) compliance with qualification requirements and rules for carrying out licensed activities in the field of environmental protection, as well as activities in respect of which a notification procedure has been established;

7) compliance by local executive bodies with the requirements of the environmental legislation of the Republic of Kazakhstan for the provision of public services in the field of environmental protection.

8) State environmental control is carried out in the following forms:

9) preventive control without visiting the subject (object);

10) preventive control with a visit to the subject (object); checks.

2.3 Collection of environmental information, its analysis and assessment

Any management begins with collecting information. Environmental information includes information and data about:

1) the state of the environment and its objects;

2) factors affecting the environment, including pollution;

3) programmatic, administrative and other measures that have or may have an impact on the environment;

4) environmental standards and environmental requirements for economic and other activities;

5) planned and implemented environmental protection measures and their financing;

6) activities that have or can have an impact on the environment, the decision-making process and the results of environmental inspections, including the calculations, analyzes and other information related to the environment considered during this process;

7) the impact of the state of the environment on the health, safety and living conditions of the population, cultural objects, buildings and structures.

Environmental information may be expressed in written, electronic, audiovisual or other form.

From the function of collecting environmental information, its analysis and assessment, two functions are distinguished as derivatives, which today have become independent: accounting of natural resources and maintaining state cadastres of natural resources; state environmental monitoring. Organization of state accounting of natural resources and maintenance of a unified system of state cadastres of natural resources. In essence, accounting data also constitutes environmental information. Accounting and registration of natural resources are a prerequisite for organizing management in this area. Features of natural resource accounting include:

a) accounting is carried out in state cadastres of natural resources, which represent a systematic collection of information on quantitative and qualitative indicators of natural resources;

b) accounting, as a rule, is carried out in kind, and not in monetary value;

c) accounting is carried out continuously, which is due to the natural variability of the state of natural resources, therefore constant updating of information is required.

2.3.1 Organization of state environmental monitoring

Government bodies organize environmental monitoring by state-owned enterprises and other entities. State environmental monitoring (monitoring of the environment and natural resources) is a comprehensive system of monitoring the state of the environment and natural resources in order to assess, forecast and control changes in their condition under the influence of natural and anthropogenic factors. The objects of state environmental monitoring are atmospheric air, land, surface and groundwater, subsoil, flora and fauna, as well as the Earth's climate and ozone layer, ecological systems, and factors affecting the environment's health.

Because of we are fully digitizing not existing IT ecosystem of ecology, we at first started from researching aspects of ecology. This work is based on the study and analysis of environmental regulation mechanisms. The following studies were carried out: study of regulations in the field of environmental protection; collection of information on the submitted reports, their forms, frequency, analysis of their information content; analysis of the use of IT technologies, the need to improve and introduce new ones

The objects of automation are the main activities of the Ministry of Ecology, Geology and Natural Resources, its territorial bodies, operators, processors, associations, organizations, non-governmental organizations, laboratories involved in the process of management, control and regulation of environmental management.

To perform tasks related to operational activities in the field of natural resource management and environmental protection, specialists at various levels carry out:

- collection, storage, processing and analysis of information about objects of negative impact on the environment, monitoring and assessment of their

impact on the environment;
- analysis of the implementation of environmental protection measures and the introduction of the best available technologies;
- preparation of documents and draft management decisions, reporting documents on the results of activities.

A unified information space provides for the presence of state standards both in solving problems of compliance with legislation and in implementing corporate requirements. As legislation and requirements change, the system will change simultaneously, and the entire GSM will operate within the framework of a unified approach to environmental management.

The common database will provide MEGPR with an information resource that contains various types of environmental information at various levels, enterprise – region – Republic of Kazakhstan. As information accumulates in the database, the value of this resource will increase, since when solving environmental management problems, it is important not only to assess the situation at the moment, but also its retrospective analysis. The unified database will contain not only final reporting figures, but also information from primary environmental accounting. This will allow, when making management decisions, to operate with both summary data and perform an analysis of the processes through which this or that indicator value was obtained.[1]

The joint presentation of spatial and attribute data on environmental objects and environmental impact assessments will provide the opportunity to:
- visual comprehensive analysis and modeling of the behavior of objects of negative impact, components of the natural environment (atmospheric air, water bodies, soils);
- obtaining consistent and visual information about ongoing events;
- possibilities of zoning the territory according to indicators of the state of environmental objects.

The functions of the subsystem are grouped into sets of tasks. The distribution of functional tasks depends on the access rights of the User role. The following users and their roles are defined for the EES:

External users authorizing via digital signature:
- authentication on the web portal via login and password;
- filing an application for public hearings;
- conducting a dialogue (asking questions, participating in public hearings) on the Portal;
- creating a personal account at your own discretion;
- editing data in a personal organization registered in the electronic portal.

Internal users:
- authentication of civil defense employees via login and password;
- checking information about the received data;
- control, monitoring, entering information, depending on the granted access rights.

System administrator:
- registration/authentication of the Software Administrator;
- user account management;

- software administration and maintenance;
- posting news content.

Operating Conditions of the Automation Object

It is envisaged to operate the EES under the following conditions:

1. Availability of constant Internet access.
2. Availability of computer and peripheral equipment.

Technical means must ensure the implementation of the functions and tasks of the system specified in this document, and must also be located in structural units. Servers act as centers for switching, storing, processing, and exchanging data and provide logical connection of the entire complex of tasks being solved. Interaction of workstations with databases is carried out using client - application server - database server technology (three-level architecture).[2]

Functional requirements for an automated system

1. Increasing the efficiency of state environmental control by automating state environmental control.
2. Obtaining current information about the owners of objects from the information systems of authorized government bodies;
3. Increasing the efficiency of the work of officials by reducing labor costs for compiling, collating the List, and registering acts in the bodies of the KP-SiSU;
4. Increasing trust on the part of subjects of state control and the Prosecutor General's Office;
5. Transparency of the List formation process, eliminating errors made during manual formation;
6. Reducing corruption risks during state environmental control.

The requirements described above include functional requirements for the relevant subsystems, as well as system-wide technical requirements for the automation platform being implemented. All requirements listed below in this document are mandatory requirements.

List of processes (subprocesses and functions)

Functionality of all proposed modules of the developed information system (Table [2.1]).

As depicted in Table 2.1, the system is based on client-server technology, which consists of server-side and client-side applications (web client). The use of this technology provides the following advantages:

- Absence of data duplication in the system by performing all computations on the server;
- More reliable data protection due to security systems installed at server storage locations;
- Centralized data usage;
- Centralized backup and archiving;
- Remote access to the system 24/7;
- Better client performance due to the absence of load on the computer's operating system.

To address computational power requirements, it is recommended to utilize BigData technology.

Table 2.1 – Functionality of all offered modules

FUNCTIONAL QUALITIES	Description
Data Input	Collection and processing of initial data on environmental pollution by emissions from various sources. Initial data can be entered manually or imported from an xls file.
Meteorological Parameters	Collection and processing of meteorological parameters (wind speed, wind direction, atmospheric pressure, temperature, etc.). Integration with the Kazhydromet system has been carried out for this module.
Calculation Module	Methodological and mathematical support of the information system. Algorithms for situation modeling and emission calculations. This module is planned for implementation in 2025. Preliminary test work has been carried out and presented in the form of video files. See Appendix.
Data Analysis	Processing information on quantitative and qualitative indicators of environmental pollution. Analysis of the degree of atmospheric air pollution, assessment of the condition of the ground layer air in the region.
Visualization Module	Visualization of the obtained results in graphical form displayed on a GIS platform in the form of layers.
Data Output	Outputting data to the user interface in the requested format (graphical, tabular) based on the obtained results.

2.3.2 Basic System Requirements

The developed system must comply with the following basic principles: **Scalability** Scalability - the system's capacity growth is determined by the hardware characteristics on which the proposed information system will operate, with the possibility of increasing the number of users and the volume of stored electronic documents by scaling the equipment's capabilities.[3]

- The system should provide independent scalability of each level (Database Server, Application Server).
- The Application Server should provide scalability by load balancing.
- The system should allow for easy scaling both geographically (providing the ability to connect remote workstations) and to expand data storage capabilities and increase performance.

Integration Integration - the system should consist of integrated modules built on the basis of standard customizable software suites.

Reliability Reliability - the system should provide data backup, system reboot after failure and emergency situations without loss of logical database integrity, procedures to support data processing integrity after system failures or other unplanned downtime, logical verification of input data. [4]

- The system should operate around the clock without interruption, except for cases of preventive or other maintenance work related to the system's operation.
- Adding new users within the specified requirement should not affect the system's performance or functionality.

Information Security Information Security - the system should use mechanisms to automate access restriction mode for reading and/or editing individual modules, documents, and form fields (parts), which will minimize the risk of incorrect use or abuse through the following measures [5]:

- Access to the system should be granted only after user identification using a password, rights division, audit support (logging) of each login or password guessing attempt, automatic user disconnection after a certain period of inactivity and when exiting the application program.
- The rigor of identification and authentication procedures for all subjects (users) accessing resources depends on the degree of information confidentiality. Subject identification and authentication should be supported based on the subject's name (login) and password verification.
- The system should support audit procedures for subjects accessing corresponding secure resources.
- The system should support functionality for logging actions of all users (addition, editing, deletion) [6].

2.3.3 General Description of Processes

2.3.3.1 Automation of Subjects and Objects Accounting

Within this automation, a database of controlled subjects and objects will be created. The accounting card will contain all information related to the object, including conducted checks. The created database of objects will serve as the basis for further automation.[7]

2.3.3.2 Automation of Risk Management System

Functions of automatic risk assessment of an object based on approved risk assessment criteria.[8]

2.3.3.3 Automation of Inspections and Preventive Control with Site Visits

Automatic generation of a project plan for preventive control with site visits based on data on the risk level of objects and inspections conducted during the previous period. Possibility of manual adjustment of the plan: adding or removing inspections, changing deadlines. For each planned preventive control, there is the possibility of generating a checklist, including a list of requirements for the activities of the inspected objects established by the legislation of the Republic of Kazakhstan, notification of the upcoming inspection, and an act on changing the composition of the inspection commission. The Act on the Appointment of the Inspection specifies the deadlines, inspection period, preventive control with a visit to the subject of control.[9]

2.3.3.4 Automation of Inspection Results Accounting

Electronic journal of inspection results, which will allow viewing the list of generated inspection result acts, forming an inspection result act. Ability to create an administrative case and attach documents.[10]

2.3.3.5 Automation of Interaction with Government Databases and Information Systems

Electronic interaction with government databases "Legal Entities" and "Individuals" will allow obtaining the most up-to-date registration data on subjects.[11]

2.3.3.6 Coordination Process of Inspections with the Committee on Legal Statistics and Special Records

The process of coordinating inspections with the Committee on Legal Statistics and Special Records will be fully automated through web services, without paper duplication. Documents will be signed with an electronic digital signature, providing legal significance.

2.3.4 Automated Information System EEP

Information support of the EEP is built on common principles, in a functionally unified concept, and based on the use of common instrumental and system resources for implementation.

Information support is intended to provide reliable, timely, and complete information necessary for various categories of users and for solving applied tasks in accordance with its intended purpose. Information support is a set of information processes, methods, and tools for organizing and transforming information necessary for the functioning of the system.

Information support consists of data used as directories, lists, codes, and classifiers, data generated as a result of the EEP's operation, as well as data entering the system for processing.[12]

The main tasks solved by information support are:

- Systematization of accumulated and processed information;
- Ensuring the rational organization of the logical and physical structure of data in input and output arrays;
- Processing of unified standardized forms of documents, references, messages, and implementation of a unified formalized data representation in them.

The information support includes the following information:

- Classification and coding system, providing for the maintenance of classifiers, codes, and directories to order and systematize the objects of the database, which are entered and modified.
- Information base, which should include external information for the EEP (input and output information and internal information - internal representation of problem-oriented data and normative-reference information).
- Standardized documentation system;
- Normative-reference information;
- Regulatory and methodological documents.

Information should be entered into the EEP system once at its origin and used multiple times as needed with user access rights taken into account.

For the preservation of information in emergencies and failures, the system administrator's software tools provide:

- The ability to fully or partially restore the program as a result of failure situations;
- Availability of a duplication system on backup storage devices with subsequent recovery.

The system should be compatible with related systems.

Authorization should be standardized through data that can confirm the identity of the end user of the application.

PostgreSQL should be used for databases managed by the ORM system.

Automatic system backup should occur every day.

ORM should be used for data control and storage, with backup and dump for data recovery.

2.3.5 General Requirements

The architecture of the System must be built using client-server technology:

- Database server (responsible for storing and providing data);
- Application server (responsible for the main business logic of the System);
- Web server (responsible for organizing interaction between clients and the System);
- Web browser (responsible for preparing and sending requests to the server and displaying the results of request processing).

The software and technical means of the System must comply with Internet standards and support data transmission via the HTTPS protocol. The system architecture must adhere to the following basic principles:

2.3.5.1 Modularity

Due to the diversity of functions performed, the system must consist of separate modules interacting with each other, built on the basis of configuring standard sets of software that implement system functions.

2.3.5.2 Scalability

The system architecture must have seamlessly scalable software constructs, allowing the connection of new modules (components) without any editing of existing system source code.

2.3.5.3 Fault Tolerance

It is necessary to ensure the fault tolerance of the System while servicing an unlimited number of users. The system's availability must be at the level of 99.5

- Clustering of the BackEnd application - allows, due to some redundancy of computing resources, to ensure the system's operation during peak loads and in case of failure of cluster servers, providing almost instantaneous response to a malfunction;
- Cold standby - backup computing resources are connected in case of failure of the main computing resources.

subsubsection

It is necessary to ensure the possibility of integrating the System with external systems within both already automated and new processes without changing the core architecture of the System, but only by increasing additional functionality.

The possibility of integrating the System's directories with external data sources must be provided. In particular, directories of the organizational and staff structure and correspondents' directories. The ability to organize procedures for importing flat reference data from external systems through settings, without making changes to the system's software code, must be ensured.

Rest API and SOAP technology should be used.

2.3.6 Planning and forecasting of environmental activities

Measures for environmental protection and natural resource management are taken into account in forecast and program-target documents and materials, and are included in planning projects for the socio-economic development of the Republic of Kazakhstan, national (state) programs and concepts in various areas of environmental management. Thus, the plans for the socio-economic development of the regions provide for a section "Nature management and environmental protection".

A feature of planning and forecasting in this area is the development of special programs for the improvement of natural complexes and individual natural resources. Implementation plans must be developed for such programs and concepts.

2.4 Formation of a single data bank

The unified state system for monitoring the environment and natural resources (hereinafter referred to as the unified system) is a multi-purpose system provided by the state that unites all systems, subsystems and types of monitoring operating in the Republic of Kazakhstan, covering directly or indirectly issues of environmental protection, conservation, reproduction and use natural resources, protecting life and (or) health of people from the effects of harmful factors in the natural and anthropogenic environment, as well as the impact of climate change and the predicted impacts of climate change. Participants of the unified system are:

authorized body in the field of environmental protection;

specially authorized government bodies;

organizations authorized to carry out types of monitoring included in the structure of the unified system;

individuals and legal entities who, in accordance with the EC, are obligated to carry out industrial environmental control.

The exchange of information necessary for maintaining a unified system is carried out via electronic communications by providing access to the participants of the unified system to the data bank.

The software system allows for the accumulation, processing and storage of information on a unified methodological basis, ensuring the exchange of information between participants in a unified system.

The basis of management is primary data and their subsequent analysis. Primary data includes data obtained from the results of types of monitoring and not subjected to generalization, processing or analysis. Raw production monitoring data, including data obtained from an automated environmental emissions monitoring system, are considered primary data.

2.5 Collection of primary data

Primary data is submitted to a unified system in the form of approved periodic reporting forms. Currently, there are a number of problems regarding the processing and analysis of the incoming flow of primary data both at the level of the region and the republic as a whole. The main and most relevant of them:

- there is no tool for receiving, systematizing and processing primary data from automated systems for continuous monitoring of emissions (AMS)

- a unified system for receiving primary data in the form of quarterly reports does not provide the proper functionality for systematizing this data and drawing an analytical picture

- there is no tool for analyzing and comparing data from production control and changes in environmental quality.

2.6 Analysis of the use of IT technologies in the field of environmental regulation

The state information system in the field of environmental protection and natural resource management “Ecological Passport of the Territory of St. Petersburg” has been in operation for many years. It uses space-time binding to store and process various information.

It is intended for storing and presenting information to executive bodies of state power, local government bodies in St. Petersburg, organizations and citizens:

- about the state of the environment and natural resources nom potential of the territory;
- about the level of technogenic impact on the environment;
- about potential environmentally hazardous facilities;
- about objects subject to state environmental supervision carried out by the executive authorities of St. Petersburg;
- about natural and anthropogenic processes that pose a potential threat to human life and economic activity in a given territory;
- about territories where environmental restrictions apply to any types of economic activity.

The operator of the System is the Committee for Natural Resources Management, Environmental Protection and Environmental Safety of St. Petersburg. The basis of the System is the primary data base. As a result of processing primary data base storage objects, thematic maps, user queries and other materials are compiled to provide information in accordance with the scope of the System.

The primary data base has a block structure. All information blocks are a collection of cartographic and factual data, grouped according to thematic principle:

- atmospheric air;
- surface waters;
- subsoil;
- land resources;
- green spaces;
- specially protected natural areas;
- users of natural resources;
- reference information.

Within each information block, data is divided into information layers. Each information layer is a collection of interconnected cartographic and factual data. Map information is stored as enterprise geodatabase objects. Factual information is presented in the form of a set of logically and semantically related tables. The composition and structure of tables depend on the type of information stored. Factual information is semantically related to cartographic objects. In this case, the existence of information layers consisting only of factual information is allowed.

We provide an example representation in the primary data database of the information layer "Automated Atmospheric Air Monitoring System" (Table 2.2) to illustrate the combination of cartographic coverage and factual data with different update periods.

The system integrates primarily: - data from the federal environmental monitoring system; - data from regional environmental monitoring systems, including an automated system for monitoring the radiation situation; automated atmospheric air monitoring system; automated surface water monitoring system; automated system for monitoring groundwater and groundwater levels; system for monitoring exogenous and endogenous processes; green space monitoring system; a system for monitoring soil pollution, a system for monitoring biological pollution of waters of the Gulf of Finland; - data from state executive authorities;

Table 2.2 – Example presentation in the primary data database of the information layer

Information Layer	Storage Object	Update Period
Atmospheric Air Monitoring System	Network of atmospheric air monitoring stations	Continuous
	Characteristics of stations	Long-term
	Primary data of atmospheric air monitoring	Operational
	Averaged data of atmospheric air monitoring	Operational

- information about water bodies in St. Petersburg;
- information about the subsoil of St. Petersburg, including information from the territorial geological fund;
- information about green spaces in St. Petersburg;
- information on the nature of the distribution of contaminants in soils;
- information on environmental permits issued;
- information on planning, implementation and results of regional environmental supervision;
- information from the land cadastre and real estate cadastre;
- data from state statistical reporting.

The System identifies four main periods for updating information:

several hours – for operational information;

up to quarter – for short-term information;

1 year – for long-term information;

more than 1 year – for permanent information.

All primary data placed in the System are metrologically supported and have a single space-time reference. The System also includes models for calculating the dispersion of pollutants (pollutants) in the atmospheric air, the spread of pollutants in surface and groundwater, for calculating the fallout of pollutants from the air, for calculating the risk to public health from the level of atmospheric air pollution and necessary for carrying out model simulations. database calculations.

2.6.1 Software

ESRI ArcGIS Server 10.0 is used as a geographic information system. The system is based on a MS SQL Server database server running ESRI ArcSDE and ESRI ArcGIS Server. As part of the information exchange, a service has been developed that provides data transfer in direct access mode, which can be connected on the computer of a user with access rights using ArcCatalog. Once you connect to the service, you can use it in ArcMap to draw maps and download published data. Local user software is implemented on ArcGIS for Desktop Advanced (ArcInfo10).

2.6.2 Data formats

The system stores information about the geographic location of objects in geodata format. Geodata is stored in geodatabase, shapefile, coverage, raster image, tables in .dbf and .xls format. Data visualization is carried out using a map – a data frame. The latter is represented as a geographic window in which map layers are displayed and used. Each layer represents a specific set of data overlaid on a map.

Map layers help present information in different ways:

- as classes of discrete objects (sets of points, lines and polygons);

- as continuous surfaces, such as relief, which can be represented in different ways. For example, in the form of a set of contour lines and points with heights, or as a relief with a hillshade;

- in the form of aerial photographs or satellite images covering the extent of the map.

The twenty-year experience of St. Petersburg in creating an information system designed for storing and processing state information resources, based on the latest achievements of geographic information technologies, has no analogue. The idea that the main tool for supporting management decisions is a map, to which not only factual data is attached, but also photo and video materials, is increasingly spreading throughout the world.

I would like to believe that the experience of St. Petersburg will be used in the development of a comprehensive information system for monitoring the state of the environment on the territory of the Republic of Kazakhstan, and this system itself will be implemented using modern geoinformation technologies.

Currently, work is underway in the Republic of Kazakhstan to create geoinformation monitoring systems and digitalize document flow (public services portal).

2.7 Methods and technologies using the example of atmospheric air monitoring

An environmental air pollution monitoring system that has been developed, tested, and built in line with the wireless standard is used to track the concentrations of the primary air pollutant gases. This system measures the concentrations of CO, NO2, and SO2 gases using semiconductor sensors. The hardware unit includes a single-chip CPU, a number of air quality sensors, a GSM-Module, and a GPS-Module. The Central-Server is a cutting-edge personal computer application server with internet connection. The hardware device gathers the CO, NO2, and SO2 air pollution levels and bundles them with the GPS location, time, and date. The frame is subsequently uploaded to the GSM-Modem and sent to the main server through the wireless network.

A wireless sensor network-based SO2 monitoring system as well as any affecting environmental variables including air temperature, humidity, and wind speed are the main aims of this study. The layout will also enable the distribution of data and information that can be accessed instantly from smartphones and desktop PCs. The schematic of the system design is shown in Figure [2.1]. The newly created embedded sensors may be used to measure the temperature and humidity

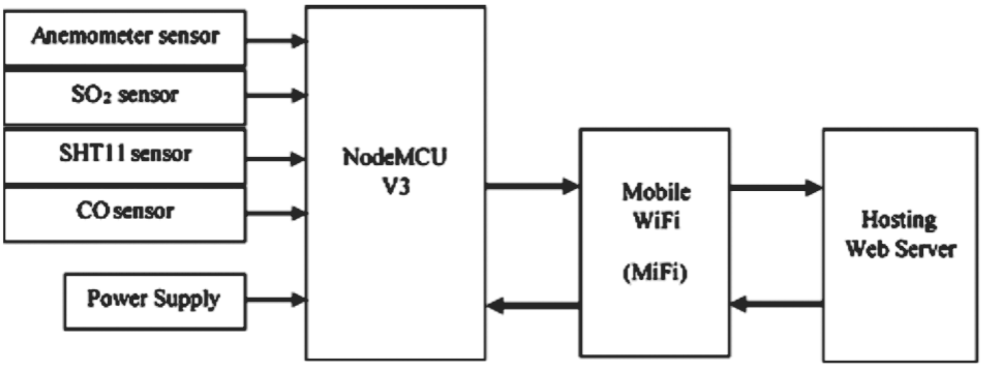

Figure 2.1 – Working scheme for monitoring ambient air quality

of the air.[13] The on-chip temperature correction and calibration, amplifier, and signal conditioning of these sensors enable direct interfacing to any microcomputer with an on-board analog to digital converter[14]

We employ the Node JS programming language and PostgreSQL database engine for the web-based user interface. Figure [2.2] depicts how the air quality monitoring system's appear on the website.[15]

2.7.1 EEP Event Logging Module

The event logging module should provide interaction with SIEM systems via the syslog protocol (RFC 5424). Monitoring of events related to information security (IS) breaches and analysis of monitoring results must be conducted.

Events related to the state of IS should be registered and violations identified by analyzing event logs, including:

- operating system event logs;
- database management system event logs;
- antivirus protection event logs;
- application software event logs;
- telecommunication equipment event logs;
- intrusion detection and prevention system event logs;
- content management system event logs.

Event logs must be stored for the period specified in the IS technical documentation, but not less than three years, and must be in operational access for at least three months.

Event logging must ensure time synchronization with the time source infrastructure. Additionally, a formalized procedure for informing and responding to IS incidents must be implemented.

Event logs must be protected from interference and unauthorized access. System administrators must not have the authority to modify, delete, or disable logs. Confidential IS requires the creation and maintenance of a backup log storage.

Events of the IS service that must be logged include:

- creation,

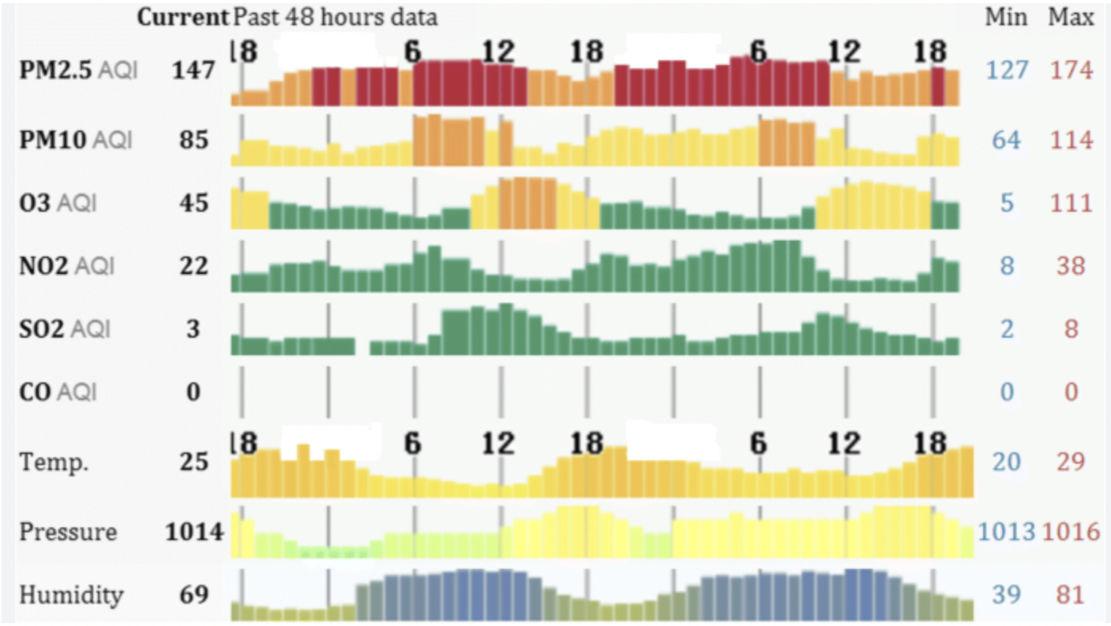

Figure 2.2 – Graph of air quality monitoring

- copying,
- moving,
- blocking,
- deletion,
- modification of local accounts and configuration files;
- failed or rejected user actions;
- user access to objects;
- user authorization (login and logout), successful and unsuccessful authorization attempts;
- all user actions in the IS object affecting IS;
- all administrator actions in the IS object.

The application log must include the following fields:
- date and time (date format: DD:MM:YYYY, time format: HH:MM:SS);
- name of the event source (service);
- account name / user ID;
- event category;
- event description;
- event level with the following attributes:
  - Alerts - requires immediate intervention;
  - Critical - critical events;
  - Errors - error messages;
  - Warning - various warnings;
  - Notifications - various important notifications;
  - Informational - informational messages;
  - Debug - for application debugging.
- client IP address;
- operation start time;

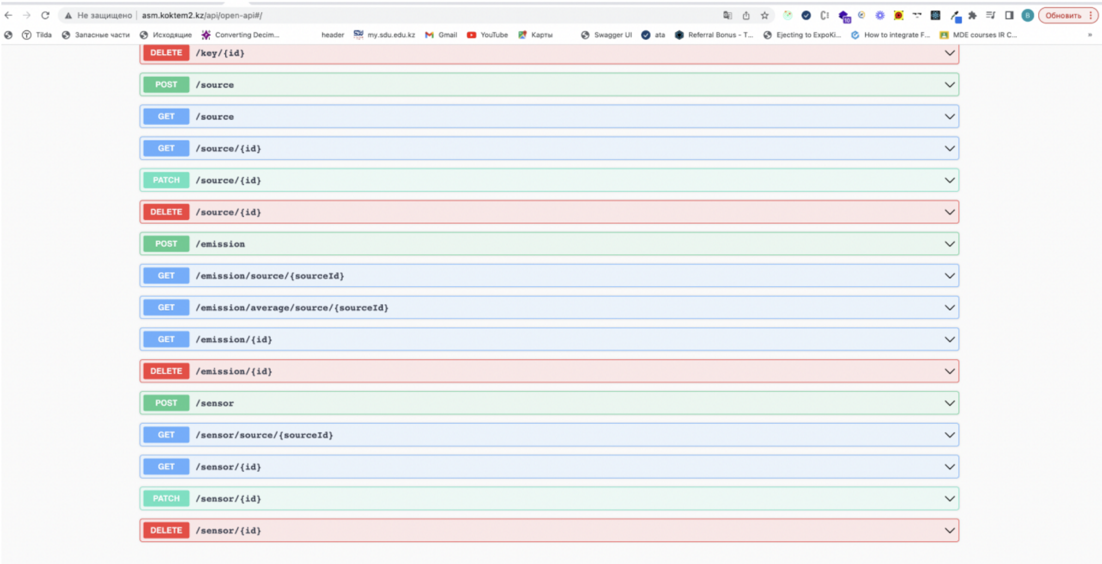

Figure 2.3 – Swagger documentation of API

- operation end time.

Requirements for the event log format:

- define the record format for events logged. Field values must be separated by delimiter characters; if the field has a long format and contains delimiter characters, use field-enclosing characters;
- the log is a text file where each line records one server access;
- UTF-8 encoding must be used for logs;
- do not record events with different data formats in one log file;
- use key-value pairs in the log;
- the folder name with log files should reflect the content;
- the IS object must support the LEEF log format.

2.8 Developed System

A web application is one of the system's components for tracking air pollution from moving cars, and it's made to provide measurement data and simulate pollution dispersion in space (fields of concentration of harmful emissions). PostgreSQL was utilized as a database. Using CSS Media Queries and Bootstrap with AntDesign UI Kit for React JS, a user-friendly and device-compatible website was created. The JavaScript library React JS is utilized for quick work because of AJAX interaction and dynamic interfaces. Google Maps was utilized with extra information, such as traffic flow rate, as geographic maps. The Google Maps APIs enable flexible customization of map interaction and viewing, the addition of graphic objects, and a wide range of data types that are suitable for the visualization of pollution. On the website, there are two maps at the moment. The first shows the measurement points with genuine measurement.

The web service structure is built after the database structure and schema are ready. There were two stages to the study's execution. The prepared database

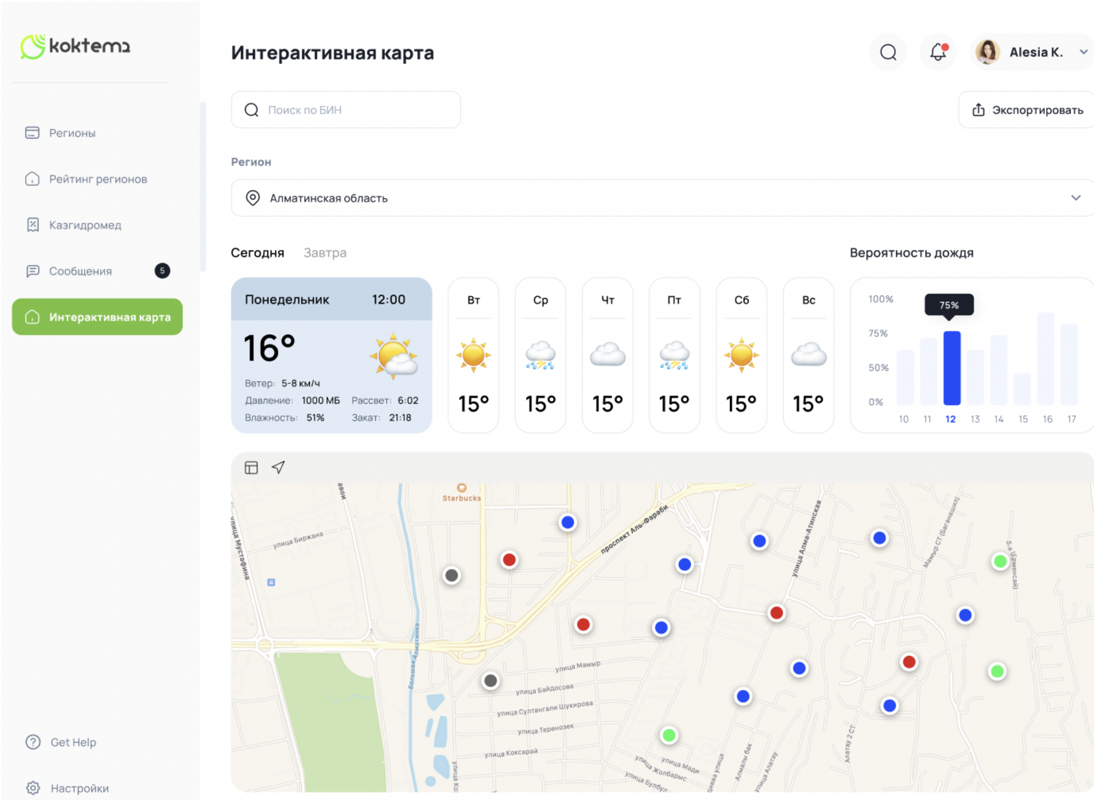

Figure 2.4 – Illustration of an auxiliary information

is presented in a secure and practical manner during the first phase, which comprises web service operations. The creation of the web service makes it possible to obtain previous AQ data safely, without wasting resources or using too much bandwidth, and it also prevents issues brought on by erroneous requests. The REST technique, one of today's most popular service-oriented architectures, was utilized to construct the web service. REST is a contemporary web architecture that is frequently used and favored because of its lightweight, simplicity, and extensibility. The REST strategy makes it simple for many development platforms to consume web service APIs, allowing consumer applications to concentrate on processing rather than data access. The fact that REST is based on the HTTP protocol makes it simple to create client-side apps as well. Compared to SOAP, a competing protocol, REST does not require the usage of a proxy or WSDL. Developers of server-based applications are now more rapid and practical adapters as a result. This research complies with the REST protocol standards in the created web service. Due to their suitability for the goals of the web service used in this study, two of REST's four pre existing request kinds have been created.

In general, there are four basic request types: GET, PUT, POST, and DELETE. In our implementation, the GET and PUT methods may be used to access the created API. Data was created for customers using the GET technique, and data from the monitoring network was loaded into the service using the PUT method. The GET method specifies 5 parameters for usage by developers of other applications. Of these four factors, three can be questioned. Figure [2.3] offers several

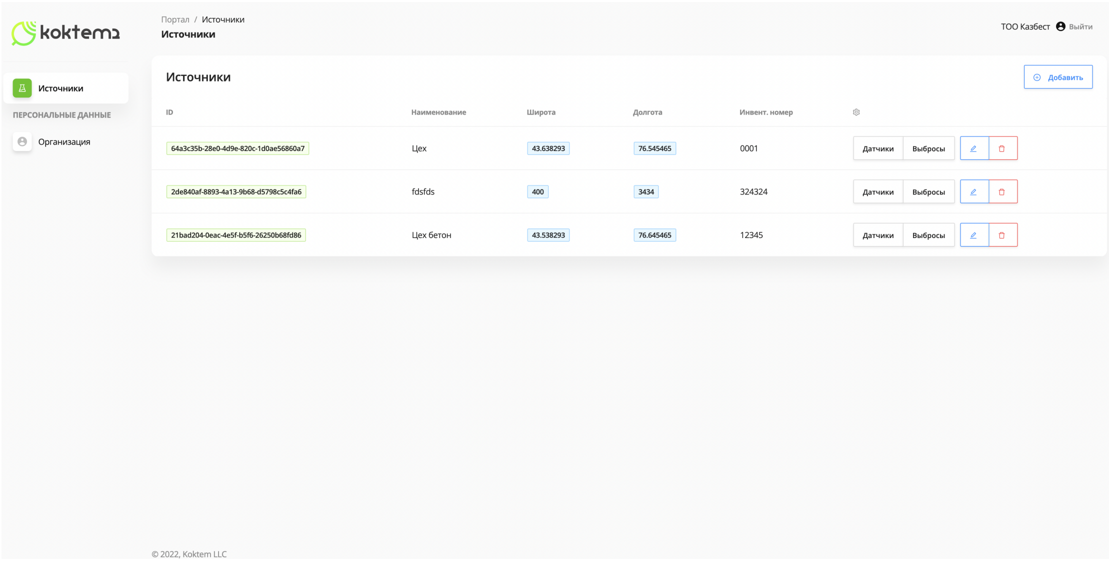

Figure 2.5 – "Sources" page of web portal

illustrations of web service request inquiries. In answer to the GET /emission/-source/sourceId query, all historical AQ data for the selected station is returned. In response to the GET /emission/average/ source/sourceId query, all historical AQ data inside the bounding box provided by averaged information is returned.

Hourly temporal granularity is integrated with the data streams from each station and each contaminant (average aggregation). Then, a verification process is used to each time series to identify any possibly inaccurate data points. For each pollutant, we determined a maximum hourly threshold value that was double the danger threshold value defined by the Peruvian air quality standard. Values beyond this level are labeled as null values to avoid distortion of the spatial modeling. The data from the numerous stations is then compiled into a matrix for each pollutant, with columns denoting stations and rows denoting aggregated hours. We employ leave-one-out crossvalidation (LOOC), a technique in which each station (column) is iteratively simulated to be an unknown point and the remaining stations are used as known points for the estimation of the unknown point values, in order to calculate an error metric between the predicted and actual values.

Another aspect of the suggested approach is short-term forecasts (6 h in advance) of target pollutant concentrations (CO, NO2, and PM2.5) at station locations. The processes for developing forecasting models, including data processing, model training, model testing, and model deployment, are described in this section. Each map has a switch at the top where you can choose whether the measure is displayed for today, this month, or this year. Additionally, visitors to the website can view more information by clicking on a particular location on the map. An illustration of an auxiliary information window is shown in Figure [2.4] and Figure [2.5], which includes additional indicators like humidity and air temperature in addition to the ones shown on the map. This window's content and presentation style are configured from the administrative section.

In addition to allowing users to view maps, the website also includes news, blogs, and pages with a wealth of useful information. There is a page with contact

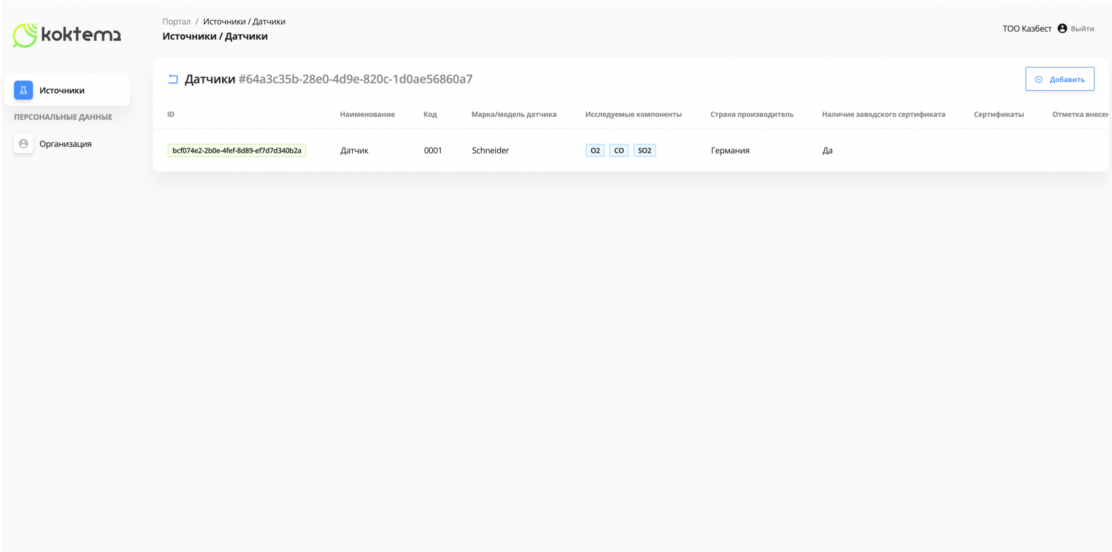

Figure 2.6 – "Sensors" page of web portal

information and a feedback form where site visitors can ask questions or submit ideas. The website also features a forum where users can discuss issues like air pollution and ways to protect the environment. Create new administrators and grant them only specific permissions using the administrative profile. The website's content can be easily edited, news can be created, and articles can be written. Controlling the display of information about air pollution is the responsibility of the administrative profile's specialized section.

"Sources" (Figure [2.5]) - All of the sites where the system gets measurements are listed in this list. It is important to provide a name, latitude, longitude, and description while amending. Given that there may be several sensors in one area, this section is designed to organize them.

"Sensors" (Figure [2.6]) - The list of all the system's sensors is available in this section. The name, code, measurements it transmits (indicators), and placement of a new sensor must all be specified when attaching it to the system. As a result of the lack of measurements or their inaccuracy, the system will automatically recognize sensor or communication issues. Additionally, because there is no rigid connection to other parts of the system, the application may continue to function normally even when they are modified, which means that while work is being done on the actual site, it won't interfere with the performance of other parts like the sensors.

Chapter 3

Testing system on real company

3.1 Big companies

Our AMS system at this moment working in some big companies of Kazakhstan. Total Dunga is one of the biggest oil production companies and subsidiary company of ”KazakhMunayGas”. ”KazAzot” is Kazahstani national manufactorer and exporter of Ammonia and Ammonium Nitrate. Also in testing mode at two big CHP as ”Atyrau TEC” and ”Stepnogorsk TEC”

3.2 Description of AMS testing system

AMS Service is the web service which collects emission data from sensors and send average data for 20 minutes to SmartBridge portal using SOAP request. The SOAP request must be signed with the E-Signature issued by PKI.GOV.KZ for the Company.

Table 3.1 – Server Information

Hostname	**Role**	**IP Address**	**OS**
KZAKT-SRVDMZ003	Integration server	172.20.29.4	Windows
KZDUN-DMZAPP007	Transition DB server	172.18.100.20	Windows
KZAKT-SRVAPP021	Application server	10.126.10.76	Windows
KZAKT-SRVSQL001	Application level SQL server	10.126.10.75	Windows

The reporting application that was written in Node.js is collecting data and making the SOAP requests. In Figure 3.1[3.1] top level of Dunga AMS system scheme created by using C4 modeling technique

As shown in Table 3.1, the system comprises several servers with different roles and IP addresses. This servers are local system architecture list of Dunga company.

3.2.1 C4 model

The C4 model is a software architecture modeling notation. It was created by developed by Simon Brown. Actually, international competitions of last couple

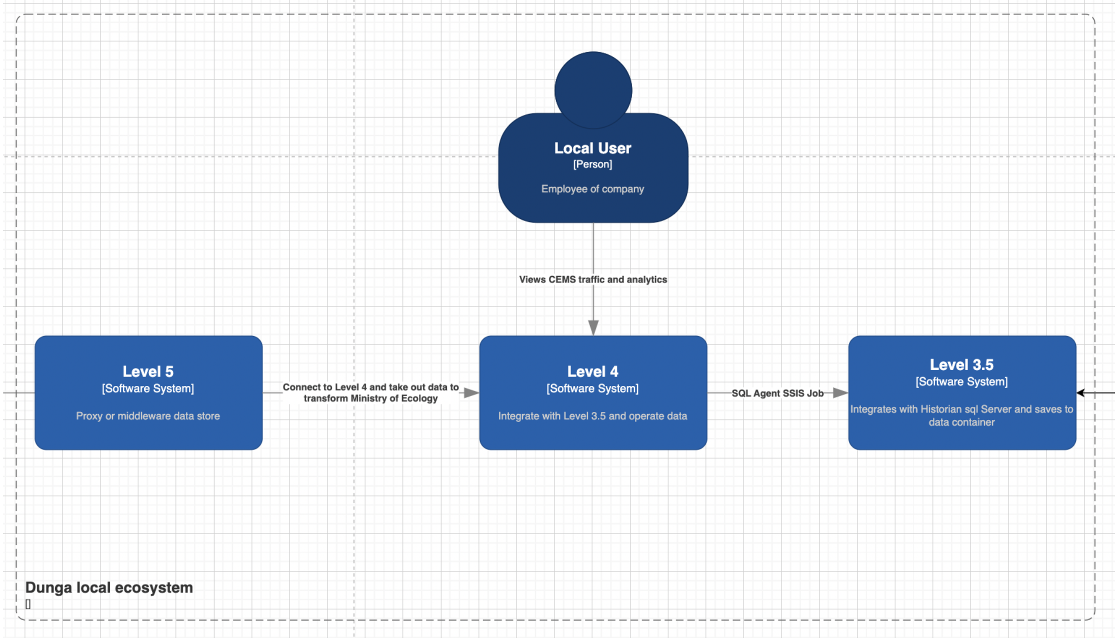

Figure 3.1 – The top level scheme of AMS service

years on area of modeling, won by people who uses C4 model notation. It's used for creating diagrams that describe system architectures.

C4 was introduced as a response to the complexity and ambiguity of traditional architectural diagrams. Simon Brown proposed this approach in the early 2010s with the goal of making architecture easier to understand for all members of a development team, including developers, testers, analysts, managers, and other stakeholders.

The name **C4** stands for:

- **Context**
- **Containers**
- **Components**
- **Code**

These four "C"s represent the four levels of abstraction used in this notation for visualizing architecture.

3.3 Data transfer architecture

The data transfer is one of main parts of architecture structure. Our developed system is nothing without source of emissions and data. There are several types of system architectures in many companies and our goal is to be very flexible to be able to collect data from all available sources. Because of it we set responsible for data transfer from emissions sources to our system nature users, so they might have own systems like SCADA or any other systems like SAT. We have availability to take out data from various of database types. Also there is a server application(back end) written on Node.js which connects to Level 4 MSSQL database and using the "Data Normalize" component converts data to required format and

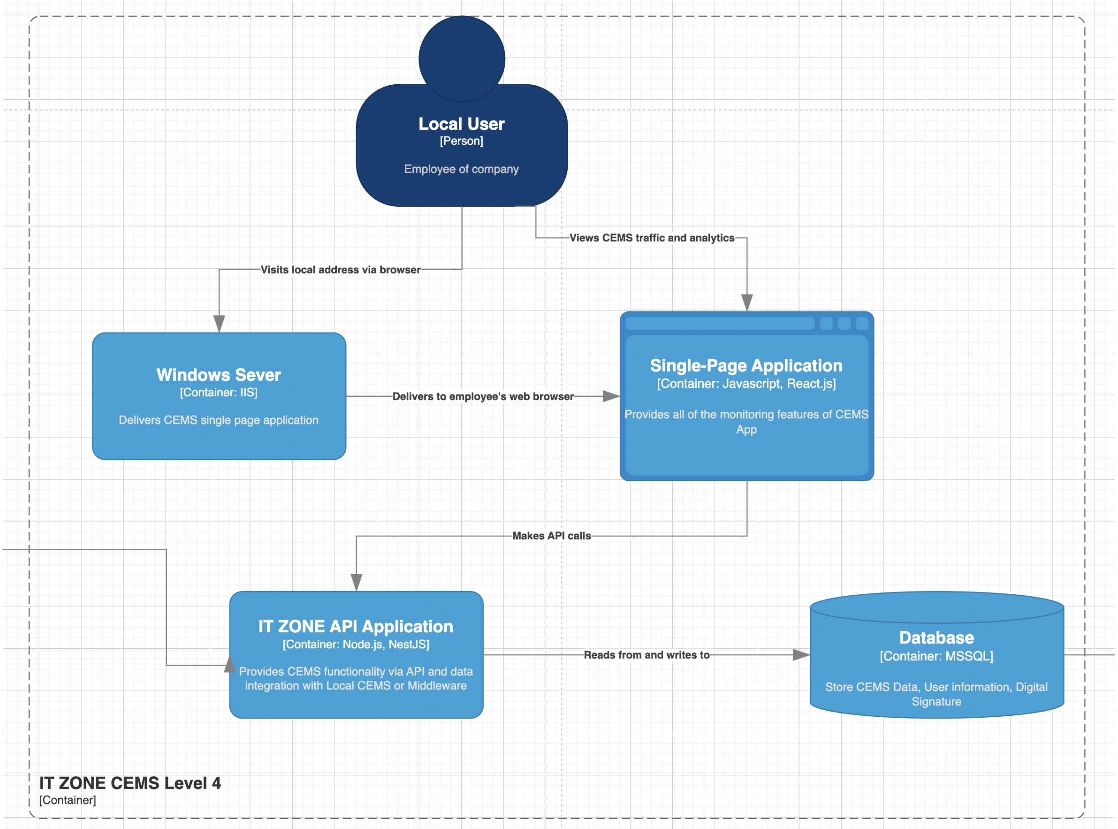

Figure 3.2 – Container level scheme of AMS service

calculates average. We can notice "Analytics component" which is used to filter and visualise data and "Monitoring Component" which is used to create directories for application functionality. And we use controllers to send data through the REST API. In the IIS located the Front side of the application written as Single Page Application using React.js which makes API calls to our backend. In both servers installed MSSQL and created databases which holds data, also enabled SQL SSIS Agency to schedule script which each 20 minutes collects data and transfer to database on Level 3.5 using Linked Servers and openquery Level 5 which is responsible for data transfer to the Ministry of Ecology CEMS. There are 3 components, "Data fetching" which is responsible for fetching required data from Level 4 using API call, "Digital Sign" component which is required for signing convert which will be sent and "Data sending" component which directly sends data to Smart Bridge using API call

3.3.1 Library Details

Applications backend was written in Node.js and frontend side in React.js External modules and libraries used in the application

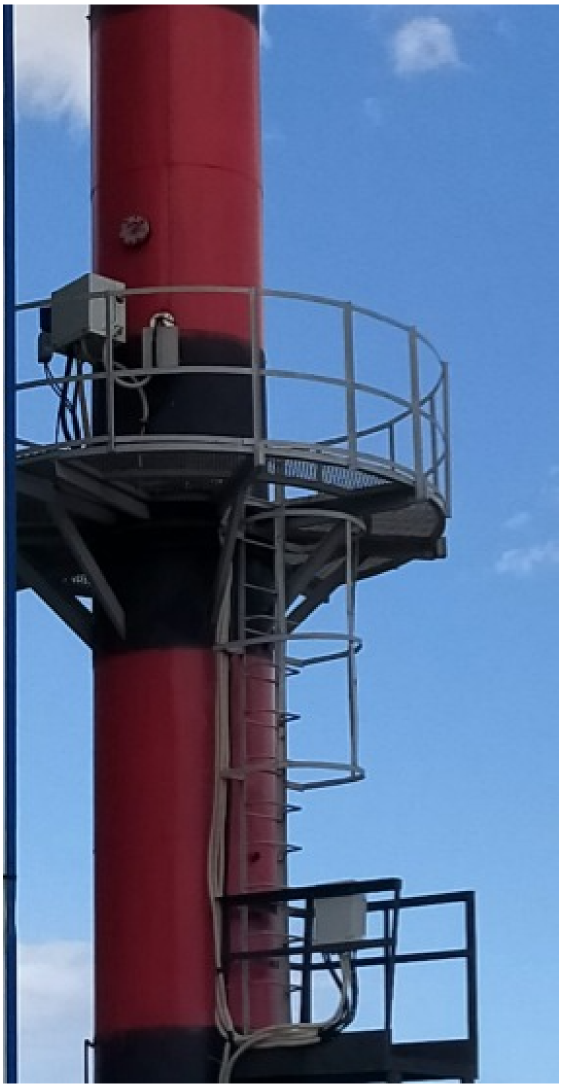

Figure 3.3 – Installation of sensors on emission tubes

3.3.1.1 Redux

Description: Redux is a predictable state container for JavaScript apps, commonly used with libraries like React or Angular for managing application state.

Purpose in Project: Redux can be used to manage the global state of your application, making it easier to manage data flow and state changes in complex applications.

3.3.1.2 Nest.js

Description: Nest.js is a progressive Node.js framework for building efficient, reliable, and scalable server-side applications. It uses TypeScript and is heavily inspired by Angular.

Purpose in Project: Nest.js can serve as the backend framework for your project, providing a structured and scalable architecture for building RESTful APIs or GraphQL servers.

3.3.1.3 MSSQL

Description: MSSQL is a database management system developed by Microsoft. It is widely used for storing and retrieving relational data.

Purpose in Project: MSSQL can serve as the database management system for your project, allowing you to store and manage data efficiently.

3.3.1.4 TypeORM

Description: TypeORM is an Object-Relational Mapping (ORM) library for TypeScript and JavaScript. It allows you to work with databases using object-oriented programming techniques.

Purpose in Project: TypeORM can be used to interact with your MSSQL database, providing a convenient way to define entities, perform database operations, and handle database migrations.

3.3.1.5 bcryptjs

Description: bcryptjs is a JavaScript library for hashing passwords using bcrypt, a password-hashing function designed to be computationally intensive and resilient against brute-force attacks.

Purpose in Project: bcryptjs can be used for securely hashing passwords before storing them in the database, enhancing the security of user authentication in your application.

3.3.1.6 PM2

Description: PM2 is a production process manager for Node.js applications. It allows you to manage and keep your application online by automatically restarting it in case of crashes or system reboots.

Purpose in Project: PM2 can be used to deploy and manage your Nest.js application in a production environment.

3.3.2 Data Transfer Troubleshooting

To troubleshoot data collection from the Historian server, follow these steps:

1. Use Microsoft Management Studio application to run queries to check if data is being received.
2. Start by running the following SQL query in Microsoft Management Studio at Level 3.5:

```sql
SELECT TOP (1000) [tag], [value], [date], [id]
FROM [dunga_asm].[dbo].[sensor_data]
ORDER BY [date] DESC;
```

This query will retrieve the latest 1000 records from the sensor data table, ordered by date in descending order. [16]

Figure 3.4 – View inside of sampling system box

As shown in Figure 3.3, the sensor which will be collecting data must be installed on emission tube with specific settings. This is very complex system of gas analysing. There are 2 types of emission analysations: heat and cold type. Difference of them that one will be heating gasses to specific degree and second one instead of takes out wetness from gasses. In Figure 3.5 and Figure 3.4 you can find sampling system of hot version. [17]

Reducing emissions of pollutants from industrial sources by defining additional requirements. The agreed draft standards do not make it possible to correctly assess the enterprise's contribution to air pollution in the region. Modeling a single object does not give a real picture; much depends on background concentrations and the location of observation posts that record these indicators. Objective: reducing dose loads for the population caused by industry.

Urban environment management (urban planning, investment, transport and energy policies), risk reduction by optimizing urban infrastructure. Increasing population resistance. Objective: improving the health of the population exposed to polluted atmospheric air.

Figure 3.5 – Sampling system

In addition to tasks related to improving the network for monitoring atmospheric air pollution, it is possible to use the results of summary calculations for tasks of computational monitoring of atmospheric air pollution for those pollutants for which methods for measuring pollutant concentrations in atmospheric air have not currently been developed.

Determination of requirements for emission source parameters to ensure optimal dispersion into the atmosphere, such as source height or degree of cover. Identification of additional activities and mitigation measures. Working with the population. As the modeling results show, the contribution from thermal power plants is less than from the private sector, and high source dispersion conditions make a smaller contribution to the concentration in the surface layer of the atmosphere. Summary calculations determine the priority of measures.

Chapter 4

Mathematical simulating model

The main goal is to scientifically and methodologically develop a Kazakh methodology for calculating the dispersion of pollutants in the atmospheric air and to develop software that can become competitive in external markets. This will be based on entirely different principles of construction and commercialization, which will allow for more efficient utilization of the potential of IT technology development to foster the economic growth of Kazakhstan and promote innovative development.

One of the features of environmental regulation is the specific characteristic that direct "management of ecology" is impossible. For instance, it is not feasible to change wind speed, humidity, or induce rain to reduce the concentrations of harmful substances in the atmosphere. Therefore, it is only possible to manage the degree of influence of anthropogenic factors and other pollutants on the environment, particularly on the atmospheric air from industrial enterprises, motor vehicles, etc.

In the context of the development of modern cities, the problem of calculating dispersion and environmental modeling is highly relevant. Project planning is the basis for regulating emissions into the environment. This provides the analytical foundation for determining where a business or other economic object can be located, causing as minimal impact on the environment as possible. In this regard, the fundamental document for calculating the dispersion of pollutants in the atmospheric air needs to be updated to modern realities and adjusted according to the recommendations of foreign colleagues.

In Kazakhstan, the dispersion of emissions is evaluated using programs such as "ERA", "Ecologist", "Atmosphere", etc. (they are based on the OND-86 methodology), while in foreign countries, the ISC program is adopted as the international standard.

Currently, there is much debate surrounding OND-86. Additionally, the new edition of the Russian regulatory document MRR-2017 still raises doubts about its relevance among foreign colleagues. Our colleagues also note that according to OND-86, changing (increasing) the height of a source by a few centimeters (a fraction of a percent) can change (increase) the ground concentration at the same point by almost 2 times or more (over 100 percent). They also note that when all parameters remain the same, increasing only the height of a point source, according

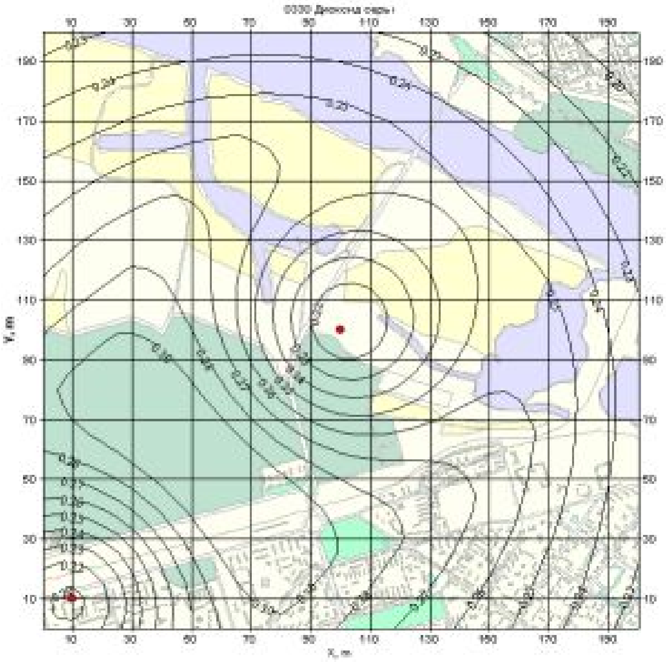

Figure 4.1 – Dispersing model on map

to OND-86, the ground concentration at a given point can increase [18].

Special attention is drawn to the fact that OND-86 is used by many countries, including Ukraine, Belarus, Kazakhstan, and others. The OND-86 regulatory document, containing an error, was used for calculations that were conducted earlier and are still being conducted daily by thousands across Kazakhstan [19]

A significant amount of work has been done by foreign colleagues to verify the correctness of OND-86 and MRR-17, for example[20]

The legislation of the Republic of Kazakhstan in the field of atmospheric air protection for environmental calculations stipulates the use of the "Method for calculating concentrations of harmful substances in atmospheric air contained in emissions from enterprises" based on the regulatory document OND-86. Based on the results of these calculations, the levels of ground-level concentrations of harmful substances are determined and then compared with the normative concentration values (MPC). When regulating emissions of harmful substances, the calculations also use the maximum ground-level concentrations calculated for specified points on the Earth's surface according to OND-86. According to the algorithm of the software "ERA", the maximum concentrations are the highest concentration values obtained by iterating wind speeds (wind directions) with a specified step within a certain range of wind speed (direction) changes. Thus, errors in the methodology for calculating ground-level concentrations at a given wind speed are transferred to calculations of maximum ground-level concentrations by iterating wind speeds with a specified step.

In the OND-86 methodology and all regulatory legal acts based on this regulatory document, formulas with interval values are used, leading to jumps of up to 120 percentage at interval boundaries. For example, when changing the height of a source in the model by 0.1 meters, if the values fall within an interval, the

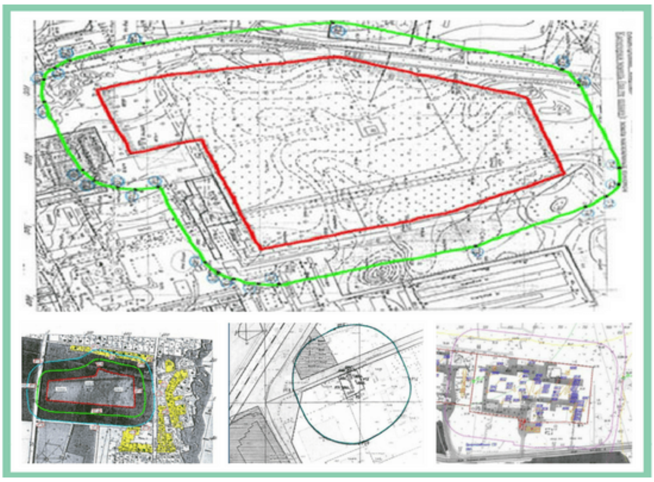

Figure 4.2 – Aermod example

pollutant concentration (PC) indicators increase by 120 percentage. In a typical scenario, an increase in the height of the stack shows a decrease in pollutant concentration. For a source with a height of 6 meters, an increase in the height of the source H from ten centimeters to 1 meter (from 2 percentage to 16 percentage) results in an increase in ground-level concentration at the same point by almost 2 times or more. This raises a practical question for environmentalists: why build smokestacks if reducing the height of the source (stack) decreases the ground-level concentration according to OND-86? This error, over 30 years old, has been carried over into Appendix 12 of the Order of the Minister of Environment and Water Resources of the Republic of Kazakhstan dated June 12, 2014, No. 221-O.[21]

Russian Federal Service for Hydrometeorology and Environmental Monitoring (Roshydromet) officials have confirmed that none of the officially recognized atmospheric pollution models in the world (dispersion of impurities in the atmosphere) from stationary point sources with a round mouth produce results with non-physical jumps (in the opposite direction) in ground-level concentration increases by more than 120 percentage when the height of the source is increased by fractions of a percent. OND-86 has a restriction on the value of the velocity U, which must be greater than 0.5 m/s. Therefore, incorrect values are obtained for maximum ground-level concentrations (obtained by iterating wind speeds in the range of 0.5 – U.[22]

Conclusion: The current OND-86 document and the second edition of the draft regulatory legal act contain errors in the formulas for calculating maximum ground-level concentrations of harmful substances. However, these corrections have not been made to Appendix 12 of Order No. 221 of 2014. An error in the current regulatory document is a threat to environmental safety. Possible underestimations of harmful substance concentrations in calculations by 100 or more (up to 150) percent have been occurring for 30 years and continue to this day. The concentration of harmful substances is underestimated in calculations according to OND-86 by 100 or more percent in several cases. Currently, emission standards are established with MPC exceedances of 100 or more percent, leading to a widespread deterioration of the environmental situation.

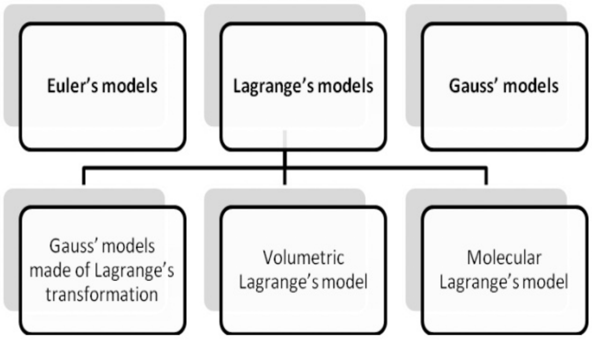

Figure 4.3 – Dispersing models

4.1 AERMOD

There are several types of models for simulating air dispersion, but there are two main methods of pollutant dispersion systems modeling: implemented with physical and mathematical models models. They both have the same starting point: broad knowledge of the lower boundary layer of the atmosphere typical. Physical models are based on execution atmospheric processes in laboratory conditions, and There is a mathematical description for mathematical models empirically based physical and mathematical processes measurement data. They are often called simulations deterministic models.[23] For correct mathematical construction it is important to classify the substances in the atmospheric air in the model. For our case in Kazakhstan Gauss's model will be best choice because of in most we have wind speed coefficient average is more than 1. It is not difficult to find a few things in it an atmosphere harmful to humans, animals or plants. The release of these pollutants causes a change in the average value specified air accumulation.[24]

So because of we stay with Gauss's model we will be working with Aermod model which has proven itself in USA, Canada, France and another countries. [25]

4.2 Open MP for digitizing algorithm

The majority of cases where these models are realized use a single-thread synchronous way of solving the problem. Instead of using a single thread, we will work with an asynchronous and parallel approach.[26]

In the development process, OpenMP was utilized to parallelize the computation and enhance performance, and SDL2 was used for visualization. Below is a code snippet demonstrating its use.

First, we include the necessary headers and define the function for saving a screenshot. This function captures the current state of the SDL window and saves it as a BMP file. The code below shows the initialization and includes necessary headers.

```
// Including necessary headers
#include "main.h"
#include <iostream>
```

```
#include "core/Window.h"
#include "models/air_pollution/AirPollutionModel.h"
#include "widgets/Canvas.h"
```

Next, we define the function `saveScreenshotBMP`. This function is responsible for capturing the current state of the SDL window and saving it as a BMP file. It reads the pixel data from the renderer and creates an SDL surface from it, which is then saved to the specified file path.

```
bool saveScreenshotBMP(std::string filepath, SDL_Window* SDLWindow, SDL_Renderer* SDLRenderer) {
  SDL_Surface* saveSurface = NULL;
  SDL_Surface* infoSurface = NULL;
  infoSurface = SDL_GetWindowSurface(SDLWindow);
  if (infoSurface == NULL) {
    std::cerr << "Failed to create info surface from window in saveScreenshotBMP(string), SDL_GetError() - " << SDL_GetError() << "\n";
  } else {
    unsigned char* pixels = new (std::nothrow) unsigned char[infoSurface->w * infoSurface->h * infoSurface->format->BytesPerPixel];
    if (pixels == 0) {
      std::cerr << "Unable to allocate memory!\n";
      return false;
    } else {if (SDL_RenderReadPixels(SDLRenderer, &infoSurface->clip_rect, infoSurface->format->format, pixels, infoSurface->w * infoSurface->format->BytesPerPixel) != 0) {
       "std::cerr << ";
       delete[] pixels;
       return false;
      }
    }
    SDL_FreeSurface(infoSurface);
    infoSurface = NULL;
  }
  return true;
}
```

The ‘saveScreenshotBMP‘ function is a critical utility for capturing the current state of the SDL window. It handles various error cases, such as failed surface creation or pixel reading, ensuring robustness. This function will be used later to save visual outputs of our simulation.

Next, we initialize the main function, where the SDL window and the air pollution model are set up. The window dimensions are defined, and the air pollution model is initialized with the window as a parameter. This setup is crucial for establishing the environment in which our simulation will run.

```
// Main function initialization
int main(int argc, char* args[]) {
  Window win(800, 400);
  AirPollutionModel model(win);
  model.initParticles();
  model.draw();
  bool quit = false;
```

```
  SDL_Event e;
  bool enter_pushed = false;
```

In this part of the main function, we set up the SDL event loop. The event loop continuously polls for events, such as key presses or window closing. Based on the event type, different actions are performed, such as updating the model, reinitializing particles, or quitting the application.

```
// Event loop setup
  win.render();
  unsigned long long cnt = 0;
  while (!quit) {
    while (SDL_PollEvent(&e) != 0) {
      if (e.type == SDL_QUIT) {
        quit = true;
      }
      if (e.key.keysym.sym == SDLK_RETURN) {
        enter_pushed = true;
      }
      if (e.key.keysym.sym == SDLK_BACKSPACE) {
        enter_pushed = false;
        model.initParticles();
        model.draw();
      }
      if (e.key.keysym.sym == SDLK_SPACE) {
        enter_pushed = false;
        model.update();
        model.draw();
      }
    }
```

Finally, the inner loop ensures that the model is constantly updated and displayed while the enter key is pressed. This segment is important for the dynamic aspect of modeling, allowing data to be updated in real time based on user input.

OpenMP uses a branch-merge parallel execution model. An OpenMP program begins as a single thread of execution, called the initial thread. When a thread encounters a parallel construct, it creates a new thread group consisting of itself and a number of additional threads, and becomes the master of the new group. All members of the new group (including the main group) execute code inside the parallel construct. There is an implicit barrier at the end of the parallel structure. After the parallel construction, only the main thread continues to execute user code. A parallel region can be nested with other parallel regions, in which each thread of the original region becomes the main thread for its group of threads. Nested regions can in turn include regions at a deeper level of nesting.

```
// Model update and rendering
    if (enter_pushed) {
      model.update();
      model.draw();
    }
    win.render();
  }
  return 0;
}
```

The main function serves as the entry point of the program, orchestrating the initialization, event handling, and continuous updating of the simulation. This structured approach ensures a responsive and interactive simulation environment.

Code above is one part which is used for visualization of our model.

Chapter 5

Conclusions and future work

5.1 Conclusions

To reduce the hazards of excessive exposure to polluted air, it is important for residents in metropolitan areas to be aware of the present and projected state of the air quality they breathe. However, given the significant costs of acquisition and upkeep, many communities may find it unaffordable to create official monitoring networks. In this article, we discussed a low-cost method of air quality monitoring that may give a person information about the quality of the air around him both now and six hours from now. The system, in this research paper, is made up of inexpensive Internet of Things (IoT) stations, AI models, and a web application that can present the anticipated information graphically (pollution maps). We evaluated the accuracy of the temporal and spatial prediction using a variety of tests. When compared to reference methodologies, the error levels were acceptable for estimating particulate matter (PM2.5). We come to the conclusion that this inexpensive approach can assist in identifying high-risk regions for exposure to airborne contaminants while also being cost-effective. Compared to numerical operational models, the deployment of the system requires less computational resources. Where there are no resources to create reference networks, it is simple to reproduce.

During the course of work, a database has been created for practical use. When scaling the project, there is an opportunity to create a whole range of data with integration into the common system, which will allow:

- Regular calculations automatically with a specified frequency (for example, every 3 hours) for forecasting changes in the quality of the atmospheric air near the area, taking into account the physiographic features of the area and current meteorological information;
- Scenario calculations for forecasting the transfer of pollutants in the atmosphere from various pollution sources located in the area;
- Calculations to assess the location of a possible pollution source that led to an abnormal increase in pollution levels at one of the observation points;
- Presentation of operational meteorological observation data, objective analysis data, and numerical forecast of meteorological variables and visualization of the distribution fields of pollutant concentrations in the atmospheric air

over the territory.

Despite the fact that the information obtained by automatic stations reflects the real state of the atmospheric air in the measurement locations, it only concerns the consequences of pollution, while its causes remain unknown. Moreover, this information shows the pollution level only at individual points and cannot provide an adequate picture of the air condition over the rest of the territory. To address these issues, mathematical modeling of pollutant dispersion in the atmospheric air is carried out, which allows assessing the degree of air pollution in a given area without conducting actual measurements. In addition, using this approach, it is possible to forecast changes in the atmospheric air condition in the long term, model various hypothetical situations (such as the construction of a new plant or even a city), and plan necessary measures to prevent air pollution. Modeling requires a comprehensive consideration of many factors, such as emission source parameters, current meteorological conditions, dispersion conditions for the given area, and properties of the modeled substances.

In conclusion, based on the experience gained in creating systems in the field of ecology, it is worth noting that the use of software in environmental research has many advantages. It simplifies data collection and processing, allows for analysis and modeling, and enables the forecasting of ecological processes. This approach allows for a deeper understanding of the environment and the development of measures for its preservation. As a result, software is an indispensable tool for modern research in the field of ecology.

In recent times, modern humans are becoming increasingly environmentally conscious. This process is facilitated by a deep understanding of the need to preserve the environment for one's own health. People are beginning to make more efforts to correct mistakes made in relation to nature and to think even more about how not to harm the environment in the future. Digital technologies, which are increasingly covering various aspects of human life, have now begun to penetrate into the ecological sphere.

At the moment, general principles and structures of global information systems have been developed to solve the problems of human health and environmental protection. However, the potential in this area far exceeds our capabilities.

5.2 Future work

To get a real picture, it is necessary not only to create an information database, but also to ensure integration with existing ones for automatic updating, for example, with data from industrial environmental control reports.

We intend to add more layers of data to the AI models in next work to give them a more local context and enhance forecasts. We want to feed the models information that is known.

I act as a junior researcher as part of a group of scientists from Al Farabi University and are participating in scientific work on the digitalization of a mathematical model. 5 month ago we won funding from the state and activly working on it.

Bibliography

[1] RD 50-34.698-90, "Methodological guidelines. Information technology. Standards complex and guidelines for automated systems. Requirements for the content of documents."

[2] GOST 19.504-79, "Unified system of program documentation. Programmer's manual. Requirements for content and design."

[3] Ministry of Digital Development, Innovation and Aerospace Industry of the Republic of Kazakhstan, "Rules for drafting and reviewing technical specifications for the creation or development of informatization objects of the 'e-government'," Order No. 143/NQ, June 2019.

[4] State Standard of the Republic of Kazakhstan, "Information Technology. Automated systems standards complex. Technical requirements for the creation of an automated system," ST RK 34.015-2002.

[5] Ministry of Digital Development, Defense and Aerospace Industry, "On approval of methods and rules for testing informatization objects of the 'e-government' and information systems classified as critical information and communication infrastructure facilities for compliance with information security requirements," Order No. 111/NQ, June 2019.

[6] Government of the Republic of Kazakhstan, "On approval of unified requirements in the field of information and communication technologies and ensuring information security," Decree No. 832, December 2016.

[7] GOST 27818-88, "Computing machinery and data processing systems. Permissible noise levels in workplaces and methods of determination."

[8] GOST 25861-83, "Computing machinery and data processing systems. Electrical and mechanical safety requirements and test methods."

[9] GOST 26329-84, "Computing machinery and data processing systems. Permissible noise levels of technical equipment and methods of determination."

[10] GOST 34.601-90, "Information technology. Automated systems standards complex. Automated systems. Stages of creation."

[11] GOST 21552-84, "Computing equipment. General technical requirements, acceptance, test methods, marking, packaging, transportation, and storage."

[12] GOST 34.201-89, "Information technology. Automated systems standards

complex. Types, completeness, and designation of documents during the creation of automated systems."

[13] "IoT-based automatic monitoring and control systems." [Online]. Available: https://www.researchgate.net/publication/351938363_IoT_Based_Automatic_Monitoring_and_Control_System

[14] "A review of automated monitoring systems in construction projects." [Online]. Available: https://www.researchgate.net/publication/294871803_A_Review_of_Automated_Monitoring_System_in_Construction_of_Road_Project

[15] "TimescaleDB: Time-series data management." [Online]. Available: https://www.timescale.com/

[16] GOST 34.603-92, "Information technology. Types of automated systems testing."

[17] GOST 12.1.012-2004, "Occupational safety standards system. Vibration safety. General requirements."

[18] "Issues with OND-86 calculations." [Online]. Available: http://www.ecotalk.ru/phorum/viewtopic.html&f=9&t=1026

[19] "Errors in the OND-86 methodology." [Online]. Available: https://blog.pravo.ru/blog/32505.html

[20] "Criticism of the OND-86 methodology." [Online]. Available: http://neftianka.ru/wpcontent/uploads/2015/04/criticism-of-the-ond-86.pdf

[21] Republic of Kazakhstan, "Environmental Code of the Republic of Kazakhstan," Code No. 400-VI, January 2021. [Online]. Available: https://adilet.zan.kz/rus/docs/K2100000400

[22] Unified Environmental Portal, Ministry of Ecology, Geology and Natural Resources of the Republic of Kazakhstan. [Online]. Available: https://company.ecogov.kz/

[23] "European Environment Agency: Air quality assessment." [Online]. Available: https://www.eea.europa.eu/publications/2599XXX/page012.html

[24] "Modeling environmental impacts." [Online]. Available: https://home.iitk.ac.in/~anubha/Modeling.pdf

[25] "Advanced modeling techniques for air quality analysis." [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1566253521000658#d1e593

[26] "Openair: Open-source tools for air quality data analysis." [Online]. Available: https://davidcarslaw.github.io/openair/