



# Collection and Processing of Data to Optimize the Monitoring of Atmospheric Air Pollution

Zh. O. Oralbekova, Z. T. Khassenova<sup>(✉)</sup>, and M. G. Zhartybayeva

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan  
zthassenova@mail.ru

**Abstract.** The system of environmental monitoring in Almaty is a collection of data without analysis and processing, without identifying sources of pollution. An urgent solution to this problem is to expand the system of environmental monitoring of the city, since the foothill zone of Almaty is characterized by extremely weak resources for self-cleaning the atmosphere. When developing an information system to optimize the monitoring of atmospheric air pollution by heavy metals, the key stage is the collection, processing and structuring of data [1,2]. The object of research is the air of the surface layer of Almaty, selected at the monitoring posts for the pollution of the air monitoring network. In the course of the research, a selection of sites for air sampling and determination of heavy metals in the samples was carried out. The received data were structured and prepared for inclusion in the created database and further complex mathematical calculations. Technical requirements and limitations to the information system are given.

**Keywords:** Ecology · Emissions · Heavy metals · Air · System · Monitoring · Database · Database management system · User

## 1 Introduction

The protection of the air environment is one of the priority directions of the policy of the Republic of Kazakhstan in the field of ecology.

Therefore, gathering and analyzing data on atmospheric air pollution is of particular relevance. The integrated environment will provide the opportunity to process, forecast, produce complex calculations and will help to avoid a large number of emissions of harmful substances into the atmosphere.

The quality of atmospheric air in the city is determined by the emissions of pollutants from the enterprises located on its territory and road transport - the main negative factor that determines the growing pollution of all environments and the concern of citizens [3]. This factor will remain the dominant one in assessing the environmental situation in Almaty.

According to the International Quality of Living Survey, Almaty occupies the 176th place among the 230 largest cities in the world in terms of environmental

quality [4]. At the same time, among these 230 cities there are also “ten million”. According to Almaty Urban Air, some air samples in the “southern capital” showed a degree of air pollution 9 times higher than normal. According to the UN Assembly on the Environment, in the death of more than 25% of children who did not live to 5 years, environmental degradation is responsible.

70% of harmful emissions into the air environment of Almaty fall to the share of transport. According to the Almaty Traffic Police Department, to date, more than 850,000 cars have been registered in the city. When burning fuel, the greatest amount of toxic impurities and heavy metals are released: sulfur oxide, aldehydes, benzopyrene, soot, lead compounds, causing severe diseases, including cancer. Also, studies have shown that the city of Almaty is contaminated mainly with heavy metals, whose high concentration adversely affects the environment.

Given the above problems, we have developed a database in the database management system MS SQL Server for data collection and processing to optimize monitoring of air pollution. In this article we will make the Classification of a priori information, development of attributes of the information under investigation and proceeding from this, we will justify the structure of the created database [5,6].

## 2 Rationale for Database Creation

The air quality management system includes state monitoring of ambient air quality, establishment of emission standards, carrying out model calculations of the level of its pollution by atmospheric emissions from industrial enterprises and vehicles. It is being carried out with the aim of developing action plans to reduce the negative impact, as well as carrying out environmental measures aimed at reducing emissions of pollutants into the atmosphere.

The database allows to integrate the available data on the quality of atmospheric air and to conduct their joint analysis [5]. It includes a primary data storage unit.

The sources of primary data are:

- results of automated quality control of atmospheric air;
- data on stationary sources of emissions of pollutants into the atmosphere;
- data on emissions of pollutants into the atmosphere from vehicles.

### 2.1 Object of Research

The object of the research was air in the surface layer of Almaty, which was selected for 2 monitoring stations for pollution of the air monitoring network: Station No. 1 st. Amangeldy/Ave. Abay and Station No. 2 - Ave. Raimbek/st. Nauryzbai batyr. For our studies, the observation data for these posts monitoring the content of heavy metals were analyzed.

The selection is made at 2 points in the city 1 time per week. The data obtained as a result of laboratory analyzes are used to construct maps of the

distribution of concentrations of pollutants - both the main pollutants (heavy metals) and a number of volatile organic compounds [7].

The database uses an internal data format that excludes the ability to import from standard dbf or csv files.

### 3 Classification of a Priori Information

The information used and generated in the process of air pollution can be divided into the following relatively independent units:

- 3.1. Information about users and their roles in the system;
- 3.2. Information on measurements of the concentration of impurities of specific heavy metals in the atmosphere;
- 3.3. Wind information.

These data can be broken down into four independent databases, or can be combined into one database. The choice depends on the DBMS, on the basis of which the client application will be developed. For example, if you use Oracle, then it is advisable to create a single database. Since at this stage we will use MS SQL Server, which allows simultaneous work with several databases. The following names for databases, tables and fields are working and can be changed. For achievement of the optimization system's objectives of monitoring of pollution of an atmospheric air with heavy metals require a solution of certain problems for optimum performance of a system. In the course of work the main processes of an information system were selected and on the basis of which the corresponding modules are created. A modular principle of creation allows ensuring optimum functioning of an information system. The main components of the system are the following modules that perform basic functions for the system operation:

- "data entry" module;
- meteoroparameters module;
- "assimilation of data" module;
- "data analysis" module;
- visualization module.
- "data output" module;

All the main tasks of this system's modules are performed on the server. A web client sends a request to the server which processes a request and replies to the client. (A web client sends server request, the server, in turn, processes a request and again sends to the client.) If there is a necessity, the server sends a request to DB. Each module has the functionality which is presented in Table 1.

#### 3.1 Information About Users and Their Roles in the System - Users

Below we describe the structure of the tables of this database (Tables 2, 3, 4 and 5).

**Table 1.** Functionality of modules

Module	Functionality of modules
“Data entry” module	Collecting, processing of input data about environmental pollution by heavy metals (concentration of impurity) of the city of Almaty. The input data can be entered and also are imported from various sources
Meteoparameters module	Collecting, processing of meteoparameters (wind speed, a direction of the wind, temperature, atmospheric pressure, etc.)
“Assimilation of data” module	Mathematical support of the environmental monitoring system of atmospheric air with heavy metals is variational algorithms for sequential data assimilation in real time
“Data analysis” module	Definition of the extent of an atmospheric air pollution, a status assessment of a ground layer air of the atmosphere of the city
Visualization module	Visualization of the received results in a graphic view with use of geoinformation technologies
“Data output” module	On the basis of the received results, data output in the form of reports on the extent of pollution of an atmospheric air of the city is made

**Table 2.** User: FD\_AuthUserType

Fields:	Data type	Field description	Field properties
id	bigint	Identifier user	Key field
name_ru	Varchar	Name user	Required (not null)
fullName_ru	Varchar	Full name of the user	Values are allowed NULL
Ord	bigint		Values are allowed NULL
Vis	Integer		Values are allowed NULL

**Table 3.** User Authentication: AuthUser

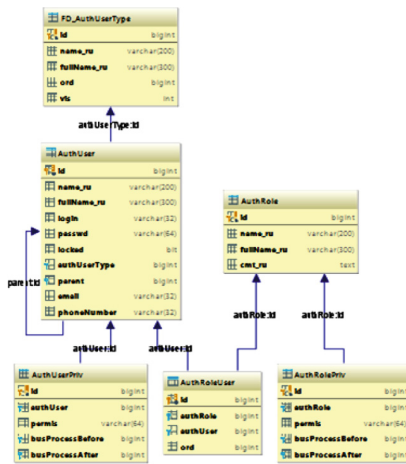
Fields	Data type	Field description	Field properties
Id	Bigint	record id	Key field
name_ru	Varchar	Name user	Required (not null)
fullName_ru	Varchar	Full name of the user	Values are allowed NULL
login	Varchar	User login	Required (not null)
passwd	Varchar	Password	Required (not null)
locked	bit	User Status (Lock)	Required (not null)
authUserType	bigint	Authorized User Type	Required (not null)
parent	bigint		Required (not null)
email	Varchar	Mail user	Required (not null)
phoneNumber	Varchar	Phone	Values are allowed NULL

**Table 4.** Authorized user role: AuthRole

Fields	Data type	Field description	Field properties
Id	Bigint	Identifier id	Key field
name_ru	Varchar	Name user	Required (not null)
fullName_ru	Varchar	Full name of the user	Values are allowed NULL
cmt_ru	text	Comments in Russian	Values are allowed NULL

**Table 5.** Authorized user role: AuthRoleUser

Fields	Data type	Field description	Field properties
id	bigint	Identifier id	Key field
authRole	bigint	The role of an authorized user	Required (not null)
authUser	bigint	The role of an authorized user	Required (not null)
ord	bigint		Values are allowed NULL
vis	bigint		Values are allowed NULL



**Fig. 1.** User unit

The FD\_BusProcessType, BussProcess, AuthRolePriv, AuthUserPriv tables are filled in the same way (Fig. 1), and contain additional information about the authorized users of the system. Figure 1 shows the relationship of the tables in the User block.

### 3.2 Information on Measurements of the Concentration of Impurities of Specific Heavy Metals in the Atmosphere

The main table in this database is a table with data on measurements of the concentration of impurities of specific heavy metals in the atmosphere - Pollution, the types of materials and their physical characteristics, which may have implications for collection and analysis. Figure 2 shows the “Pollution” block.

This database in the future will consist of two parts. One of them corresponds to standard mathematical models of media, the other to models created in the natural experiment. In the course of the work, this database should be expanded and supplemented with new data.

Below describe the structure of the created tables (Tables 6, 7 and 8).

**Table 6.** DataCollectionPoint – basic information about data collection points

Fields	Data type	Field description	Field properties
id	bigint	Data collection point identifier	Key field
name	Varchar	Data collection point name	Obligatory field
lon	float	Longitude	Obligatory field
lat	float	Latitude	Obligatory field
cmt_ru	text	Comments	optional field

Data on heavy metals and other pollutants in the atmospheric air are stored in the PollutionElements table.

**Table 7.** PollutionElements

Fields	Data type	Field description	Field properties
Id	Bigint	Records ID	Key field of the table
name	varchar	Name of pollutant	Required, foreign key
alias	varchar	Secondary name (formula)	Optional
concentration	Float	Concentration of the contaminant in the atmosphere	Required, foreign key

The Measurements table contains data on measurements at collection points.

### 3.3 Wind Information

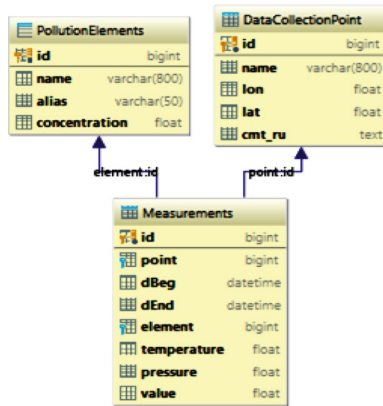
The wind is the key factor in the spread of pollutants in the ambient air, and it is advisable to create a separate data block for the wind parameters. The FD\_WindType, Wind (Tables 9 and 10) contains basic information about the characteristics of the wind.

All tables can be added to the database as needed. Five tables are designed to write system data – Dblog, dbFileStorage, sys\_jc\_GenId, DbInfo, Wax\_VerDb ModelingOthers, which will provide information about the files, the database.

To collect and process data to optimize the monitoring of air pollution, a database is created, which is shown in Fig. 3 as a relational model of the database.

**Table 8.** Measurements

Fields	Data type	Field description	Field Properties
Id	bigint	Records ID	Key field
point	bigint	data collection points	Obligatory, Key
dBeg	datetime	Start time of the conducted measurement	Obligatory
dEnd	datetime	End time of the conducted measurement	Obligatory
element	bigint	Contaminant	Obligatory, Key
temperature	float	Air temperature during measurement	Obligatory
pressure	float	Air pressure during measurement	Obligatory
value	float	Value	Obligatory



**Fig. 2.** Block “Pollution”

**Table 9.** FD.WindType

Fields	Data type	Field description	Field Properties
Id	bigint	Records ID	Key field
name_ru	varchar	Name	Obligatory, Key
fullName_ru	varchar	Full name	Obligatory
ord	bigint		values are allowed NULL
vis	bigint		values are allowed NULL





## 4 General Structure of the Monitoring System

The working process of the air pollution monitoring and surveillance system is shown in Fig. 4 and consists of three main blocks, namely: a database, a real-time sensor data reporting system from two monitoring stations and the server. Finally, the results of the previous steps can be used for applications.

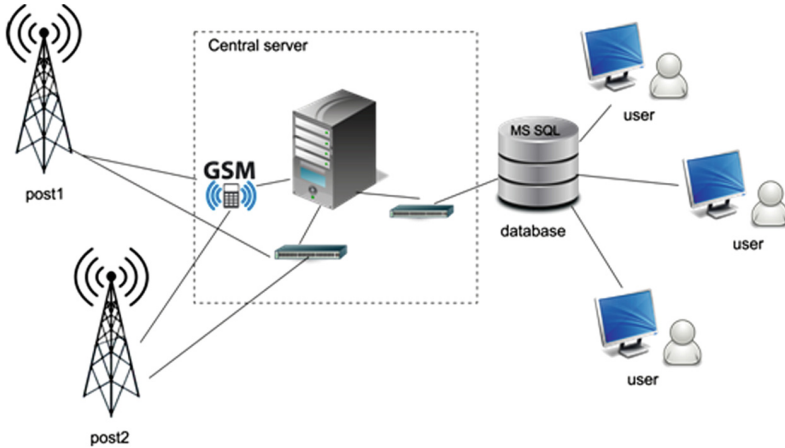


Fig. 4. General structure of the monitoring system

## 5 Conclusion

Despite all the efforts made by the Government of Almaty to improve the quality of atmospheric air in the city, air pollution remains the main environmental problem of our metropolis. The constant growth of the fleet of vehicles in combination with the growth of the number of industrial enterprises and the increase in the capacity of enterprises of the energy complex lead specialists to forecast a negative trend in the development of the ecological situation. In this connection, the use of modern technologies of spatial processing and data analysis makes it possible to significantly improve the efficiency of air quality management. The integrated approach implemented in the information system allows combining the maintenance of databases of emission sources, instrumental and calculation methods for air quality control, as well as using spatial analysis tools in organizing information support for making managerial decisions of the urban planning. The adoption of environmentally friendly management decisions is the basis for sustainable development of the city and the region as a whole.

The work was supported by the Ministry of Education and Science of Republic of Kazakhstan, under the Grant Development of a new information system and database to optimize monitoring of atmospheric air pollution by heavy metals (contract number 132, 12 March, 2018).

## References

1. Zhumagulov, B.T., Temirbekov, N.M., Turganbaev, E.M., Denisova, N.F., Bitimbayev, I.T.: MInformation system of ecological monitoring and application of variational algorithms. *Bull. Nat. Eng. Acad. Repub. Kaz.* **1**, 10–18 (2013)
2. Rakhmutellina, S.Zh.: The forecasting subsystem of the air pollution monitoring information system. *Search* **2**, 243–250 (2010)
3. Wang, Y.: Using propensity score to adjust for unmeasured confounders in small area studies of environmental exposures and health. Imperial College. London Department of Epidemiology and Biostatistics (2017)
4. Quality of living city ranking [Electronic resource]. <https://mobilityexchange.mercer.com/Insights/quality-of-living-rankings>. Accessed 10 Apr 2018
5. Ainsworth, D., Butcher, S., Sternberg, M., Knottenbelt, W.: Computational approaches for metagenomic analysis of high-throughput sequencing data. Ph.D. Degree of Imperial College London and the Diploma of Imperial College Imperial College London, Department of Life Sciences (2014)
6. Plusquina, M., Guid, F., Polidorod, S., Vermeulen, R., Raaschou-Nielsen, O., Campanella, G.: DNA methylation and exposure to ambient air pollution in two prospective cohorts. *Environ. Int.* **108**, 127–136 (2017)
7. Bind, M.A., et al.: Beyond the mean: quantile regression to explore the association of air pollution with gene-specific methylation in the normative aging study. *Environ. Health Perspect.* **123**(8), 759–765 (2015)