

# Analysis of Availability of Data Sets Necessary for Decision Making in Air Quality Assessment

**B Kh Sanzhapov, N M Rashevskiy, A A Sinitsyn**

Department of Mathematic and Information Technologies, Volgograd State Technical University, 28, Lenin avenue, Volgograd 400005, Russia

E-mail: rashevsky.n@gmail.com

**Abstract.** The article analyzes the problems in obtaining the environmental information data sets necessary for decision making in the field of air quality assessment. The authors of this article describe their experience in search for data required to run the WRF and Calpuff systems to estimate the air quality in the city of Volgograd, the Russian Federation. The problems the authors encountered and other researchers may encounter are indicated. A conceptual scheme of the resource that will help to increase the speed of searching for necessary information and also inform users about the rules for use of a particular resource is suggested.

## 1. Introduction

In recent years, the problem of modeling air pollution has not lost its relevance. Building of these models is necessary for their subsequent use in decision support systems. Each year, these systems are being improved, and for their development, more accurate and complex models are needed. Development of more advanced models requires large data sets. At the initial stages of research, the problem of choosing the methodology of decision-making is solved. One of the main criteria for choosing a methodology is the availability of the necessary data. Finding this data can take a very long time, because researchers may not be aware of the availability of data that can help them. Searching for data can lead to the fact that it will not be possible to use one or the other methodology of decision making due to poor data quality: some data may be missing, data may contain a pass, etc. In this article, authors analyze the problems that researchers may encounter in process of searching for and accessing data, and also offer a concept of a resource that would help users to obtain information about available data for a given territory and time period in a short time.

## 2. Experience in data access

Carrying out research in the field of air quality, in particular modeling the dispersion of emissions in atmospheric air [1-10] requires archival data sets. These archives can be small in size, for example, only the parameters of emission sources are required to use the OND-86 model [11]. More modern systems need more data, in addition to information on the parameters of emission sources, meteorological data and terrain data are required [12].

Most of the archival data is available for public access, another part of the data can be obtained in specialized organizations by request.



However, the search for data sources can take considerable time, which significantly reduces the speed of research. Let's consider the main problems of access to the necessary data by the example of simulation of atmospheric air pollution in the city of Volgograd with WRF / Calmet / Calpuff software package [13-15].

The following data sets are required for this simulation system:

- Archive of static surface data for the WRF system. The archive can be downloaded for free on the resource [16].
- Archive of weather data in GRIB format. Access can be obtained free of charge on the resource [17] after registration.
- Information about the terrain and land use for the Calmet program [18].
- Data from radiosondes on the state of the upper atmosphere for the Calmet program [19].
- Data from surface weather stations for the Calmet program [20].
- Parameters of the emission sources for Calpuff system.

The first problem encountered by the authors is the lack of necessary data at Russian resources. Parts of necessary data can be obtained from government agencies, but access to them is done on a fee-based basis, moreover, the data is provided in format that is not supported by Calpuff system.

In addition to the problem of directly searching for data sources, there is a problem of using resources on which the archive data is presented. Each resource has its own features for registering and presenting data. It should be noted that for WRF and Calmet, similar data is required, but it is not enough to take them from a single source, because the data formats do not match.

Another problem in the search for data is the gaps in this data. An important task is to find such a time periods when there is enough data to carry out modeling. In carrying out our research, we received information from various sources. In almost all of the archives there were periods with missing data: in the request for values of concentrations of pollutants at stationary posts in the city of Volgograd, there were no data for some months, because the equipment was not verified; in the requested meteorological data from one source there was no air humidity, and in the data from the other there were missing values of the wind speeds at different altitudes.

Another problem is the use of special terminology related to different scientific fields. An example is the use of various map projections: Universal Transverse Mercator, Lambert Conformal Conic, Tangential Transverse Mercator, etc. Here, you may have difficulties with converting one type of projection to another. It takes time to learn all the notations and rules. It should also be noted that virtually all data resources are using English language. This creates great difficulties for the Russian users - they may misuse the terminology and it is difficult for them to navigate the resource in a foreign language, which directly complicates the work with this resource.

Solving the problems described above takes a long time. To search for necessary information, it is required to study a lot of scientific articles, as well as sites where the use of necessary modeling systems is discussed.

A conceptual scheme of the resource that will help to increase the speed of searching for the necessary information, and also inform users about the rules for using a particular resource is suggested in the next paragraph

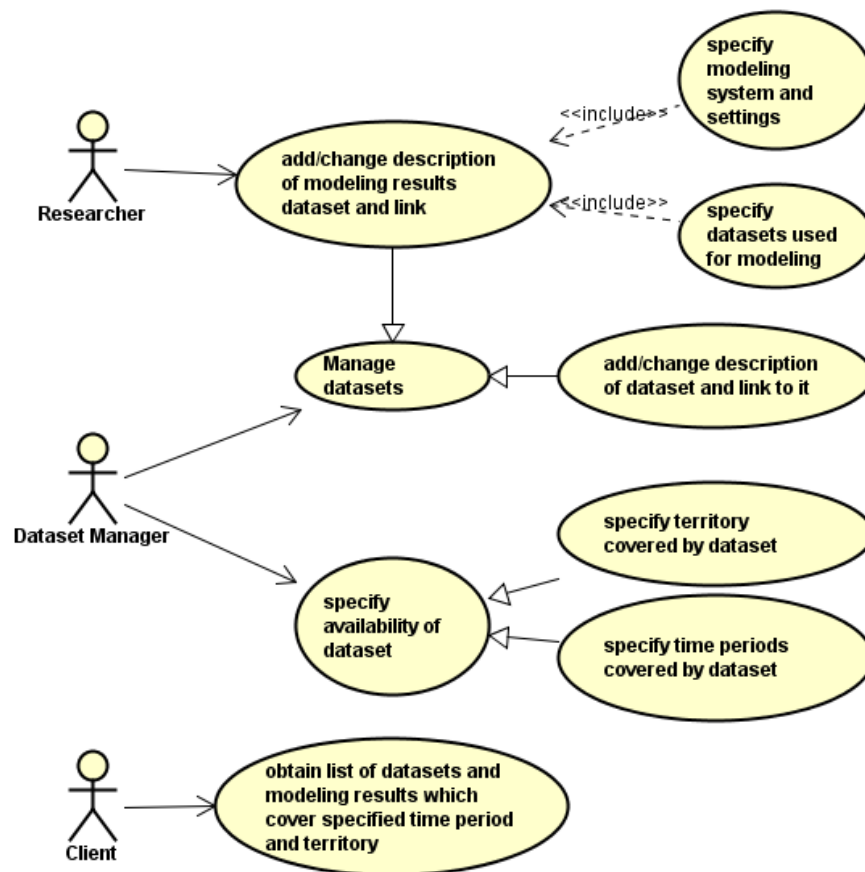
### **3. Experience in data access**

The authors suggest a concept of a resource that will allow accumulating knowledges about the availability and quality of data sets that can be used in decision making in the field of air quality assessment. UML use case diagram of resource is shown in figure 1.

The Dataset Manager is a user who often works with data sets from one or more sources. He can enter a brief description of the data set and the source, from which it is obtained, in his native language. Also, the Dataset Manager can give a link to the video instruction for using this resource. The main task of Dataset manager is to specify the time periods and territories for which one or another data set is available.

The Researcher is a user who uses software systems for modeling atmospheric processes and air quality. The result of the Researcher's work is a certain data sets characterizing a certain territory in a certain period of time. The Researcher can add to the system a reference and description of the obtained data sets so that other users can use this information for decision making. The Researcher also needs to specify the system he used, the system run settings, and the data sets used in the simulation.

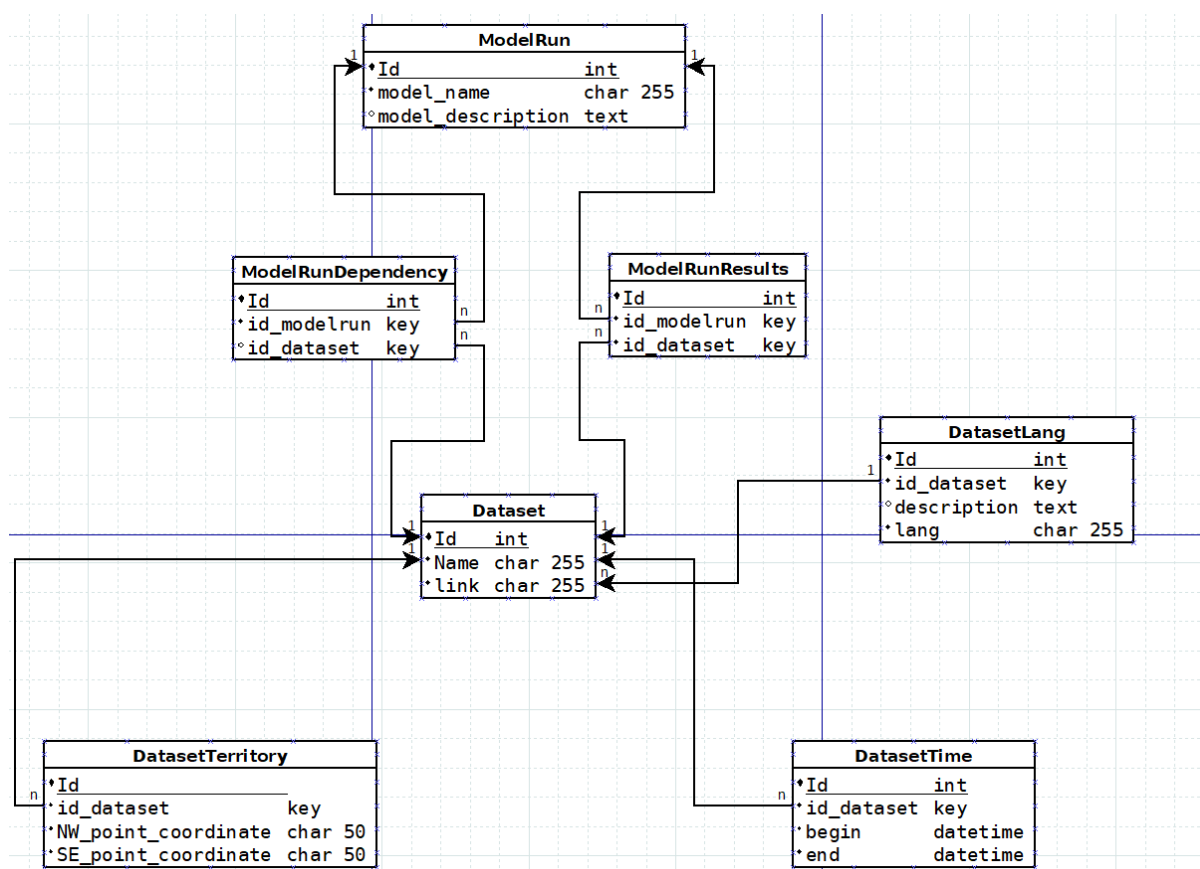
The User of the system can select a domain on the map, set a time period and obtain a list of data sets that would be accessible to him.



**Figure 1.** Use Case UML diagram of the proposed resource

The scheme of the resource database is shown in figure 2. The resource proposed by the authors is not complicated for development, because do not assume a heavy load on the server. The resource will not store the data sets themselves, but only descriptions and links to them, so the database size will be insignificant.

The “Dataset”, “DatasetTerritory”, and “DatasetTime” tables are the main tables of the resource. They contain information about data sets, coverage of territories and time periods, respectively. The “DatasetLang” table contains descriptions of data sets in different languages. The “ModelRun”, “ModelRunDependency”, and “ModelRunResults” tables are responsible for describing the research data sets.



**Figure 2.** Scheme of the resource database

#### 4. Conclusion

The development of the proposed resource will help researchers to avoid the need to analyse the availability of data for a specific territory and time periods. When solving tasks related to air quality assessment, researchers will be able to quickly estimate the amount of available information on the state of the environment. In addition, researchers will not have to spend time studying complex software systems if the simulation in these systems has already been carried out previously for the required area and the required period of time.

#### References

- [1] Sanzhapov B Kh, Sinitsyn A A and Rashevskiy N M 2016 Justification of the Choice of the Site for Construction Projects on the Basis of the Atmospheric Air Monitoring *Procedia Engineering: 2nd Int. Conf. on Industrial Engineering* **150** pp 1948–53
- [2] Naddeo V, Zarra T, Oliva G, Chiavola A and Vivarelli A 2016 Environmental odour impact assessment of landfill expansion scenarios: case study of borgo montello (Italy) *Chemical Engineering Transactions* **54** pp 73–78
- [3] Atabi F, Jafarigol F, Moattar F and Nouri J 2016 Comparison of AERMOD and CALPUFF models for simulating SO<sub>2</sub> concentrations in a gas refinery *Environ Monit Assess* **188** p 516
- [4] Abdul-Wahab S, Chan K, Ahmadi L and Elkamel A 2013 Impact of geophysical and meteorological conditions on the dispersion of NO<sub>2</sub> in Canada *Air Qual Atmos Health*
- [5] Holmes N S and Morawska L 2006 A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available *Atmospheric Environment* **40** pp 5902–28
- [6] Sironi S, Capelli L, Centola P, Del Rosso R and Pierucci S 2009 Odour impact assessment by

- means of dynamic olfactometry, dispersion modelling and social participation *Atmospheric Environment* **44** pp 354–60
- [7] Yu H and Stuart A L 2016 Impacts of compact growth and electric vehicles on future air quality and urban exposures may be mixed *Science of the Total Environment* **576** pp 148–58
  - [8] Ivančič M and Vončina R 2013 Modelling pm10 dispersion from road traffic and industry in ljubljana basin (Slovenia: Milan Vidmar Electric Power Research Institute)
  - [9] Sanzhapov B Kh, Sinitsyn A A and Rashevskiy N M 2016 Presentation of the results of atmospheric air monitoring in substitution systems for solving problems of life support of the city *Izvestiya VSTU Ser Current problems of management, computer technology and computer science in technical systems* **185** pp 40–4
  - [10] Sanzhapov B Kh, Sinitsyn A A and Rashevskiy N M 2016 Approaches to informational training on the results of monitoring the atmospheric air barrier *Bulletin of the Volgograd State University of Architecture and Civil Engineering TS Ser. Construction and architecture* **63** pp 166–77
  - [11] 1987 *OND-86. The procedure for calculating the concentrations in the air of harmful substances contained in the emissions of enterprises* (St.-Petersburg: Gidrometeoizdat)
  - [12] Sanzhapov B Kh, Sinitsyn A A and Rashevskiy N M 2017 Exclusion of the integration of fragmented programs WRF and Calpuff for simulating the distribution of barriers in the atmosphere of the city of Volgograd *Izvestiya VSTU Ser Current problems of control, computer technology and informatics in technical systems* **196** pp 46–9
  - [13] Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Duda M G, Huang X Y, Wang W and Powers J G 2008 A Description of the Advanced Research WRF Version 3 *NCAR Tech. Note NCAR/TN-475+STR* **113**
  - [14] *Exponent Inc. CALMET UsersGuide* Retrieved from [http://www.src.com/calpuff/download/CALMET\\_UsersGuide.pdf](http://www.src.com/calpuff/download/CALMET_UsersGuide.pdf)
  - [15] *United States Environmental Protection Agency. CALPUFF Modeling System* Retrieved from [http://www3.epa.gov/ttn/scram/dispersion\\_prefrec.htm#calpuff](http://www3.epa.gov/ttn/scram/dispersion_prefrec.htm#calpuff)
  - [16] *The Weather Research & Forecasting Model. WPS geographical input data sets* Retrieved from [http://www2.mmm.ucar.edu/wrf/users/download/get\\_sources\\_wps\\_geog.html](http://www2.mmm.ucar.edu/wrf/users/download/get_sources_wps_geog.html)
  - [17] *National Center for Atmospheric Research. NCAR's Research data archive* Retrieved from <http://rda.ucar.edu/>
  - [18] *U.S. Department of the Interior. EarthExplorer* Retrieved from <http://earthexplorer.usgs.gov/>.
  - [19] *NOAA . NOAA/ESRL Radiosonde Database* Retrieved from <http://www.esrl.noaa.gov/raobs/>
  - [20] *NOAA. NCDC/NOAA* Retrieved from <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/2013/>