

Formation of High-Quality Solutions In The Construction of Aeromonitoring Data Banks

Hlib Nekrasov (✉ g.nekrasov@hse.ru)

National Research University Higher School of Economics <https://orcid.org/0000-0001-7062-1070>

Alexander Vostrikov

1National Research University Higher School of Economics

Ekaterina Prokofeva

1National Research University Higher School of Economics

Nashon Adero

Freiberg University of Mining and Technology: Technische Universitat Bergakademie

Research

Keywords: geoecological monitoring, spatial data, geospatial modeling, automatic data extraction, analytical processing of a databank.

Posted Date: August 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-741562/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: This article discusses the approach to the implementation of the project for the extraction and the methodology of preliminary processing of the obtained data with the aim of centralized accumulation for collective multipurpose use of the databank on the example of carbon dioxide emissions into the atmosphere by air transport for a given territory. It should be noted that on the basis of morphological analysis, processing, as well as the classification of spatial objects of the geodatabase and additional information, it is subsequently possible to organize, for example, a system of geoecological monitoring.

Methods: At the fundamental level, the research used integration and process-based approaches, the method of extrapolation, expert methods of evaluation, random selection and analytical comparisons, a set of methods of spatial analysis based on various instruments and sources. In this study are used of open standards OGC, web, GIS technologies and the Internet for the formation, processing and storage of spatial data, their unambiguous geolocation, the implementation of territorial selections and visualization of results.

Results: The set of data, which was organized according to the proposed and defined rules, made it possible to assess the structural processing of geospatial data, and to prepare a visual representation of the impact of aviation on the environmental situation over the designated geographic area.

Conclusions: The transport industry was chosen as the object of research, but this solution can also be successfully applied to other logistics and industrial areas. During the implementation of the project, the analysis of the subject area was carried out, the architecture of the future prototype of the databank was designed, the accumulated data from the sources was structured, and a database was selected for storing them, taking into account the provision of high availability and ensuring stable operation under high loads. For the convenience of displaying data, an interactive visualization tool with a convenient and friendly user interface has been developed.

Background

The Russian Federation, in accordance with the trends of developed world countries, plans to introduce a tax on carbon dioxide emissions into the atmosphere [1]. In this regard, it is necessary to monitor the situation with regard to emissions into the atmosphere, as well as propose ways to minimize them. Within the framework of such a task, it is extremely important to have a solution to ensure the structural processing of geospatial data by automating the process of collecting and analyzing spatial data, as well as reflecting the results on an interactive map.

The transport makes a significant contribution to the health and well-being of the country's economy. In the transport sector, commercial aviation has emerged as the fastest, safest and most progressive form of transport in just over a century. The global economy benefits greatly from the ability to move people and products around the world – quickly and safely. Today, more than 3 billion people, or almost half of

the world's population, use the services of airlines around the world [2]. At the same time, aviation must be environmentally sustainable, function harmoniously under the constraints imposed by the need for clean air and water, limited noise impacts and an acceptable climate.

Aviation today affects the environment in many ways: people living near airports are exposed to aircraft noise; streams, rivers and wetlands may be exposed to pollutants discharged into storm runoff from airports; and aircraft engines emit pollutants into the atmosphere. Aircraft engines produce exhaust gas that contains about 70% carbon dioxide (CO₂) and about 30% water vapor (H₂O) [3]. Less than one percent of exhaust gases are composed of pollutants such as nitrogen oxides (NO_x), sulfur oxides (SO_x), carbon monoxide (CO), partially burnt or unburned hydrocarbons (HC), particulate matter (PM), and other trace elements. Typically, about 10% of aircraft pollutant emissions are emitted close to the surface of the earth (less than 1 kilometer above ground level), the remaining 90% of aircraft emissions are emitted at altitudes greater than 1 kilometer. The pollutants CO and HC are the exceptions to this rule as they are generated when aircraft engines operate at their lowest combustion efficiency (while the wheels are on the ground), thus separating them by about 30% below 1 kilometer and 70% above. 1 kilometer [4].

Table 1
Airlines in the world

Region	Number of airlines
CIS countries + RF	58 + 57
Western Europe	136
Eastern Europe	40
Asia	247
North America	159
South America	69
Australia and Oceania	43
Africa	113

It is worth noting that, as shown in Table 1, there are about 57 airlines operating in the Russian Federation [5] and over the territory of the state, there are about 4,000 flights a day, which every day make a passenger and cargo transportations, which in turn leads to large emissions of carbon dioxide into the atmosphere. So, according to Professor Ulrich Schumann [6] when each kilogram of aviation kerosene is burned, about 3.15 kilograms of carbon dioxide are released into the atmosphere. It is also worth noting that each type of aircraft damages the environment in different ways, depending on its technical characteristics.

Today, there are a number of services that perform calculations with constrained input parameters. An example of such services is MyClimate service from Lufthansa [7] and an online calculator from Aeroflot

[8]. These services give the user an idea of the carbon dioxide emissions into the Earth's atmosphere. However, there are a number of disadvantages, both tools present information for a single flight, and not for a cluster of flights. Each service has a number of assumptions, for example, in the service from MyClimate, the calculation of the flight between the given points is formed according to the average statistical data of the aircraft, thereby there is an error relative to the aircraft model. This problem was solved in the Aeroflot Company, adding the choice of the aircraft model. But in either case, this is a simplified formula for calculating CO₂ emissions from IATA [9], as well as the use in the second case of a limited number of aircraft models only those that are currently used by the company.

The general problem of these services, first of all, is not in the methodology of theoretical calculations of carbon dioxide emissions into the atmosphere, but in a better presentation of data and the impossibility of generating reports on air pollution for a certain time interval. It is also worth noting the inflexibility of the proposed services in terms of using a databank for analytical and monitoring structures, such as the tax service, scientists studying atmospheric phenomena, consulting groups and ordinary Internet users. Thus, the latter needs and will demand for a new information-analytical tool devoid of the above disadvantages.

Methods

Correctly laid down architecture of any system avoids many problems, especially such as horizontal and vertical expansion, the addition of new functions, and minimization of operating errors.

One of the important stages in the design of an information-analytical tool is the definition of a set of components. So, to implement high-quality solutions for building an aeromonitoring databank, one should use the following minimum set of components and nodes:

- Web crawlers. These components are responsible for collecting data from third-party resources that do not have an API to interact with them.
- Data and custom queries processing server. The name of the resource speaks for itself, this server processes the data received from web crawlers and places them in the database, and also processes user requests stored in the database, bringing them to the form that is necessary for display on the client side, thereby saving network traffic without transferring redundant data over the network, and reducing the load on the computing resources of the client.
- Server messaging through the queue. This node of the system allows you to centralize the collection, transmission and processing of a large number of messages in continuous information flows, as well as to store this large data in a kind of intermediate storage, without worrying about the risks of their loss and system performance.
- Server(s) databases. The heart of the system itself provides data storage. Responsible for the consistency and availability of data.

Among the features of the above architecture (Fig. 1) the following ones can be distinguished:

no need to use additional software on the client side, which allows automatically implement a multi-platform client side;

the ability to connect an almost unlimited number of clients;

with a single storage location and a database management system, the minimum requirements for maintaining data integrity are met; and

regarding the amount of data - the architecture of web systems does not have significant restrictions.

To ensure the filling of the databank with the help of web crawlers, first of all, it is necessary to highlight the main entities necessary for building a model, these entities will allow to keep track of daily flights on various routes, calculate the amount of emissions for each flight based on the type of aircraft, as well as view statistics in the context airlines.

In addition to the data necessary for accounting for flights and calculating the amount of emissions, it is correct to organize the maintenance of geostructural data in relation to pollution. This data is required to be displayed to the user on the website.

In this case, we use air corridors of flights rather than static geo-squares as the basis for displaying data on a future visualization map. Since flights are strictly limited to a specific flight corridor, it can be assumed that all aircraft on the same flight have the same flight path. In fact, the flight path may differ, but these differences are not significant, they are leveled by the scale of the emission dispersion, and remain within the air corridor of the flight. Therefore, the following entities can be distinguished:

- flight departure and arrival points
- flight number
- flight coordinates
- predicted pollution

Taking into account the above presented structure of entities and the need to quickly provide information to the client, in this case, you should use several DBMS.

On the one hand, the document-oriented MongoDB DBMS should be used as a DBMS that will store operational data obtained in real time using web crawlers [10–11], since it uses JSON-like documents in the data storage schema. It is also beneficial as a convenient tool for displaying data during web development, in particular, within a JavaScript-oriented stack. Thus, the structure of the composition of the entity's geodata entities that will be stored in such a DBMS must first be reformatted into JSON format, which will describe the flight trajectory of the reference path and contain information about all flights and flights along this path, as well as take into account information about the pollution amount for each flight.

On the other hand, the relational database PostgreSQL should be used as a long-term data archive. The latter also has its advantages as it supports the appropriate format and allows you to store spatial data.

As a preprocessing of data, in particular, for calculating the carbon footprint of an aircraft flight in the implementation of the prototype, a formula was established that is regulated by best practices used in the aviation industry, according to the existing methodologies of the International Civil Aviation Organization (ICAO) and the International Air Transport Association (IATA).

$$E = \frac{ax^2+bx+c}{S+PLF} * (1 - CF) * CW * (EF * M + P) + AF * x + A,$$

where:

- E - the amount of CO₂ emissions per passenger, measured in kilograms;
- x - the flight distance, which is defined as the sum of the aircraft flight segments $GCD(x [n]) = GCD(x_1, x_2, x_3 \dots x_n)$, measured in kilometers;
- S - the average number of seats, common for all cabin classes;
- PLF - occupancy rate of passenger seats, measured as a percentage of occupied seats in relation to empty seats;
- CF - load factor, is the percentage of unoccupied commercial load of the vessel;
- CW - weight coefficient of the cabin class, measured as a percentage, represents the ratio of the weight of seats of different classes;
- EF – CO₂ emissions when fuel is burned by an aircraft, measured in kilograms of fuel consumed;
- M - the multiplier takes into account potential effects not related to CO₂, such effects are the fuel consumption of the engines.
- P - the amount of CO₂ emissions to start the aircraft, measured in kilograms of fuel consumed;
- AF - aircraft weight is measured in kilograms;
- A - CO₂ emissions from airport infrastructure, measured in kilograms.

The $ax^2 + bx + c$ part is a non-linear approximation of the function $f(x) + LTO$, where LTO is the fuel consumption during landing and takeoff, including taxi to the runway, measured in kilograms. The long distance is defined as short $x < 1500$ km, and long distance - $x > 2500$ km. A linear interpolation is used between them.

For data visualization, a site has been developed for the architecture shown in Fig. 2.

It is based on a self-developed site created on the basis of HTML 5, CSS markup language and additional JavaScript modules. To process custom queries and display results on an interactive map, a REST API was written in Flask, which allows you to link queries to MongoDB, PostgreSQL databases and the visual part using the Leaflet library. Apache Kafka message broker is used to load and unload data from REST API content to databases and vice versa [12]. This tool acts as a smooth and convenient exchange of

messages between microservices, which in turn ensures the stable operation of data streams in real time, and the reliability of their receipt by the requested party [13].

Results

In the course of applying the provided method and collecting qualitative data, as a prototype, about 57 publicly available resources were analyzed. For the analysis of spatial data [14], open resources were used, from which “raw” information was extracted by automation, which was subsequently post-processed.

As a result, the obtained data are displayed on an interactive map, with the help of which it is possible to assess the level of environmental pollution, as shown in Fig. 3.

The developed interactive map allows displaying data for each completed flight over the territory of the Russian Federation, taking into account the fuel consumed and the limits from minimum to maximum, according to which it is possible to predict emissions from an aircraft, as shown in Fig. 4.

Methods for plotting by coordinates with the possibility of scaling for quantifying data have been implemented, as shown in Fig. 5.

A filter has also been added that allows you to display certain data falling into the selected area, and display the results for aircraft that hit the specified area, and CO₂ emissions for them, Fig. 6.

There are three options for displaying by filters:

- data output according to the specified time;
- data output according to the specified airports;
- data output according to the specified time and airports.

Also on all filters presented there is a display of the total amount of CO₂ emissions and other flight characteristics, such as:

the number of all aircraft;

list of names of all aircraft;

total guaranteed emissions CO₂;

total estimated emissions CO₂.

For all future output filters under consideration, all data on CO₂ emissions, as well as other characteristics of flights / routes are recalculated on the server according to the request from the Front-end.

Discussion

The works on implemented tool continue to modernize the site, as previously indicated in Figure 2, and fill it with new functionality. It is planned to implement an API to provide data to the potential users. In development, a part of the site that will allow obtaining statistical data on airlines and their emissions in relation to time, with the possibility of the further printing reports for the end user. The databanks of this kind will be useful for at least 4 target groups:

1. **Tax services.** They can use such data as reports related to the terrain and the amount of air emissions in a given area with a list of flights that have made these emissions.
2. **Scientists.** It will allow to study atmospheric phenomena, as well as their influence on the processes of the earth. For example, visualization of data on 3D maps with the ability to display aircraft trajectories relative to their flight altitudes and the implementation of some filters will allow comparative terrain assessments with the ability to generate reports.
3. **Consulting agencies.** Thanks to the fixation of emissions relative to the named flight of the aircraft, and the latter contains a link to the company, it is possible to develop a statistical output of data for consulting companies by air carrier companies in the future. These statistics will display in real time the influence of the companies on an environmental pollution, the model range of aircraft of various airlines, and form a rating for the current period and a dynamic rating for a selected period in terms of carbon dioxide emissions from airlines. It will also allow companies to revise their fleet to improve the quality indicators of carbon dioxide emissions.
4. **Any interested user.** For his needs, he will be able to independently conduct analytics on airlines and their impact on environmental pollution in relation to the area of interest to him using a certain filter or request.

In the perspectives of adding information to an interactive map for the European countries, in order to increase the information content on CO₂ from air transport and the desire to reduce potential air emissions

Conclusion

The implemented software for the structural processing of geospatial data by automating the process of collecting and analyzing spatial data and their types / attributes, as well as visualizing the results and reflecting them on an interactive map made it possible to receive a daily environmental summary of air pollution [15]. The server part provides round-the-clock service availability, connects data from the database with the web application. As a user interface, a site with an interactive map is presented, with the ability to view statistics on carbon dioxide emissions into the atmosphere, for a certain period of time along the corridors of aircraft flights, and filters are also implemented that allow displaying statistics

according to their environmental indicators/standards over the territory of Russia [16]. Some of these features include:

- the ability to create layers of information that can be shown or hidden with one click of a button;
- adding text boxes to data points that appear when clicked to give a short summary or description;
- zoom functions that allow users to focus either on the details of a specific region or get a quick overview of a wider area;
- data can be quickly updated, and these updates are made transparent to the users; and
- points on the map can be associated with external supporting documents such as images, videos, or graphics.

All software modules have gone through the stages of the life cycle: design, creation, testing, in the near future it is planned to put the developed interactive map in the public domain with the implemented API for interacting with the service, enriching the interactive map with data from different regions of Eurasia.

Abbreviations

OGC: Open Geospatial Consortium; GIS: Geographic Information System; CIS: Commonwealth of Independent States; RF: Russian Federation; IATA: International Air Transport Association; ICAO: International Civil Aviation Organization; DBMS: Database Management System; JSON: JavaScript Object Notation; REST API: REpresentational State Transfer Application Programming Interface; HTML: HyperText Markup Language, CSS: Cascading Style Sheets.

Declarations

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Funding

The publication was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2020 – 2021 (grant № 20-04-033) and by the Russian Academic Excellence Project "5-100".

Availability of data and materials

Data can be shared upon request.

Authors' contributions

All authors have made substantial contributions to the conception and design, or acquisition of data, or analysis and interpretation of data; NHA and PEN have been involved in drafting the manuscript or revising it critically for important intellectual content; and VAV and NA have given final approval of the version to be published. Each author has participated sufficiently in the work to take public responsibility for appropriate portions of the content. All authors read and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Evolution of Russian Climate Policy. from the Kyoto Protocol to the Paris Agreement - URL: <https://www.cairn.info/revue-l-europe-en-formation-2016-2-page-39.htm> (accessed on 11 June 2021).
2. Airlines around the world transported a record number of passengers. - URL: <https://rg.ru/2018/09/09/aviakompanii-po-vsemu-miru-perevezli-rekordnoe-chislo-passazhirov.html>(accessed on 11 June 2021).
3. Environmental problems in aviation and ways to solve them. - URL: <https://cyberleninka.ru/article/n/ekologicheskie-problemy-v-aviatsii-i-puti-ih-resheniya/viewer> (accessed on 11 June 2021).
4. The impact of aviation on the environment and measures to mitigate the negative impact - URL. http://method.meteorf.ru/publ/tr/tr365/ivan_a.pdf (accessed on 11 June 2021).
5. Russian Airlines - URL. <https://www.airlines-inform.ru/russia/> (accessed on 11 June 2021).
6. How aircraft exhaust fumes affect the climate. - URL: <https://www.dw.com/ru/%D0%BA%D0%B0%D0%BA-%D0%B2%D1%8B%D1%85%D0%BB%D0%BE%D0%BF%D0%BD%D1%8B%D0%B5-%D0%B3%D0%B0%D0%B7%D1%8B-%D1%81%D0%B0%D0%BC%D0%BE%D0%BB%D0%B5%D1%82%D0%BE%D0%B2-%D0%B2%D0%BB%D0%B8%D1%8F%D1%8E%D1%82-%D0%BD%D0%B0-%D0%BA%D0%BB%D0%B8%D0%BC%D0%B0%D1%82/a-15759056> (accessed on 11 June 2021).
7. Offset your flight emissions! - URL. https://co2.myclimate.org/en/flight_calculators/new (accessed on 11 June 2021).

8. Aeroflot Calculator CO₂ - URL. https://www.aeroflot.ru/ru-ru/about/calculator_CO2 (accessed on 11 June 2021).
9. Methodology for calculating the amount of CO₂ emissions - URL. https://www.myclimate.org/fileadmin/user_upload/myclimate_-home/01_Information/01_About_myclimate/09_Calculation_principles/Documents/myclimate-flight-calculator-documentation_EN.pdf (accessed on 11 June 2021).
10. MongoDB Documentation - URL: <https://docs.mongodb.com> (accessed on 11. June 2021).
11. Scrapy (web crawlers) documentation - URL. <https://docs.scrapy.org/en/latest> (accessed on 11 June 2021).
12. Kafka documentation - URL. <https://kafka.apache.org/documentation> (accessed on 11 June 2021).
13. Prokofeva EN, Vostrikov AV, Nekrasov HA, Redin DE Design of Navigation Satellite Systems in Effective Geomonitoring for Mining IOP Conference Series: Earth and Environmental Sciencethis link is disabled, 2021, 720(1), 012107 Building of effective management structures based on the digital technologies in the mining industry. A case study of UC RUSAL.
14. Nekrasov HA, Polivoda DE, Prokofeva EN The Geospatial Data Mining Concept Using Scrapping Technology Moscow Workshop on Electronic and Networking Technologies, MWENT 2020 - Proceedingsthis link is disabled, 2020, 9067406.
15. Nekrasov GA, Polivoda DE, Prokofeva EN Sources of spatial data in advancing technologies of subsoil use Mining Informational and Analytical Bulletinthis link is disabled, 2020, (5), стр. 164–176.
16. Adero NJ, Drebenstedt C, Prokofeva EN, Vostrikov AV Spatial data and technologies for geomonitoring of land use under aspect of mineral resource sector development Eurasian Mining, 2020, 2020(1), стр. 69–74.

Figures

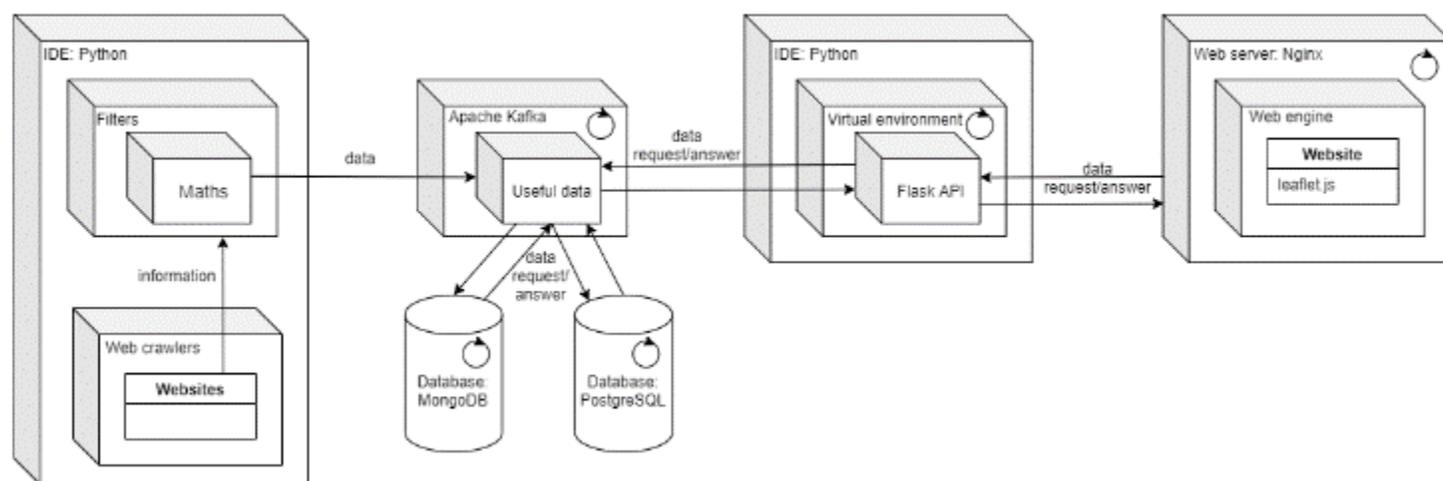


Figure 1

The architecture of the designed tool.

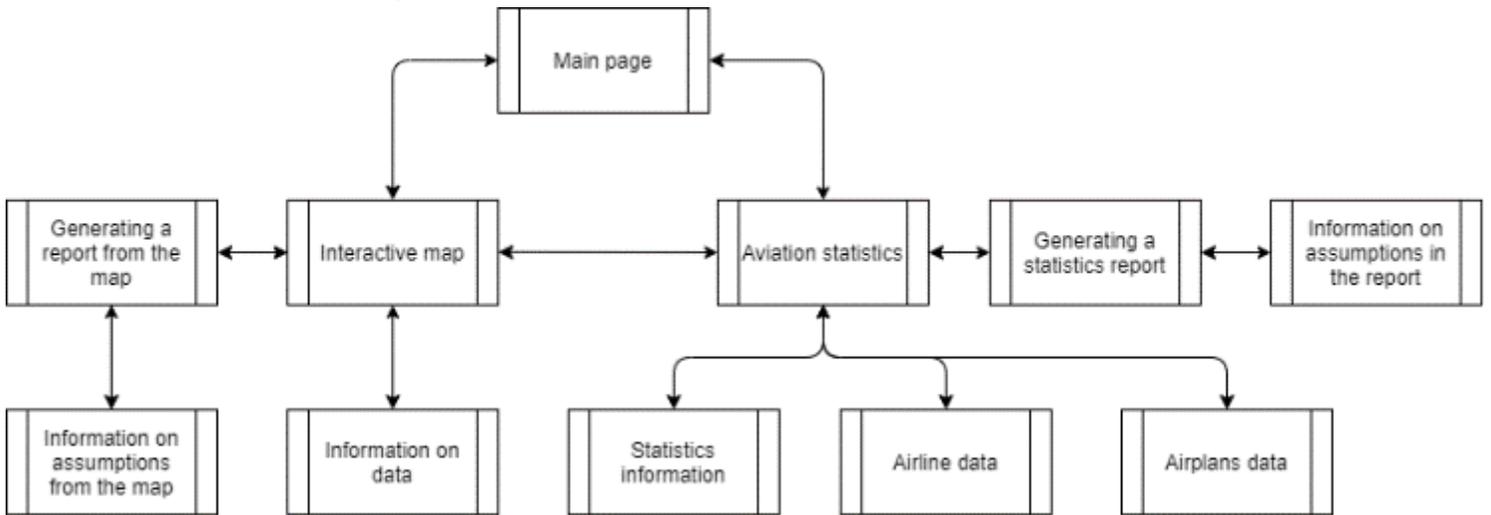


Figure 2

Site architecture.

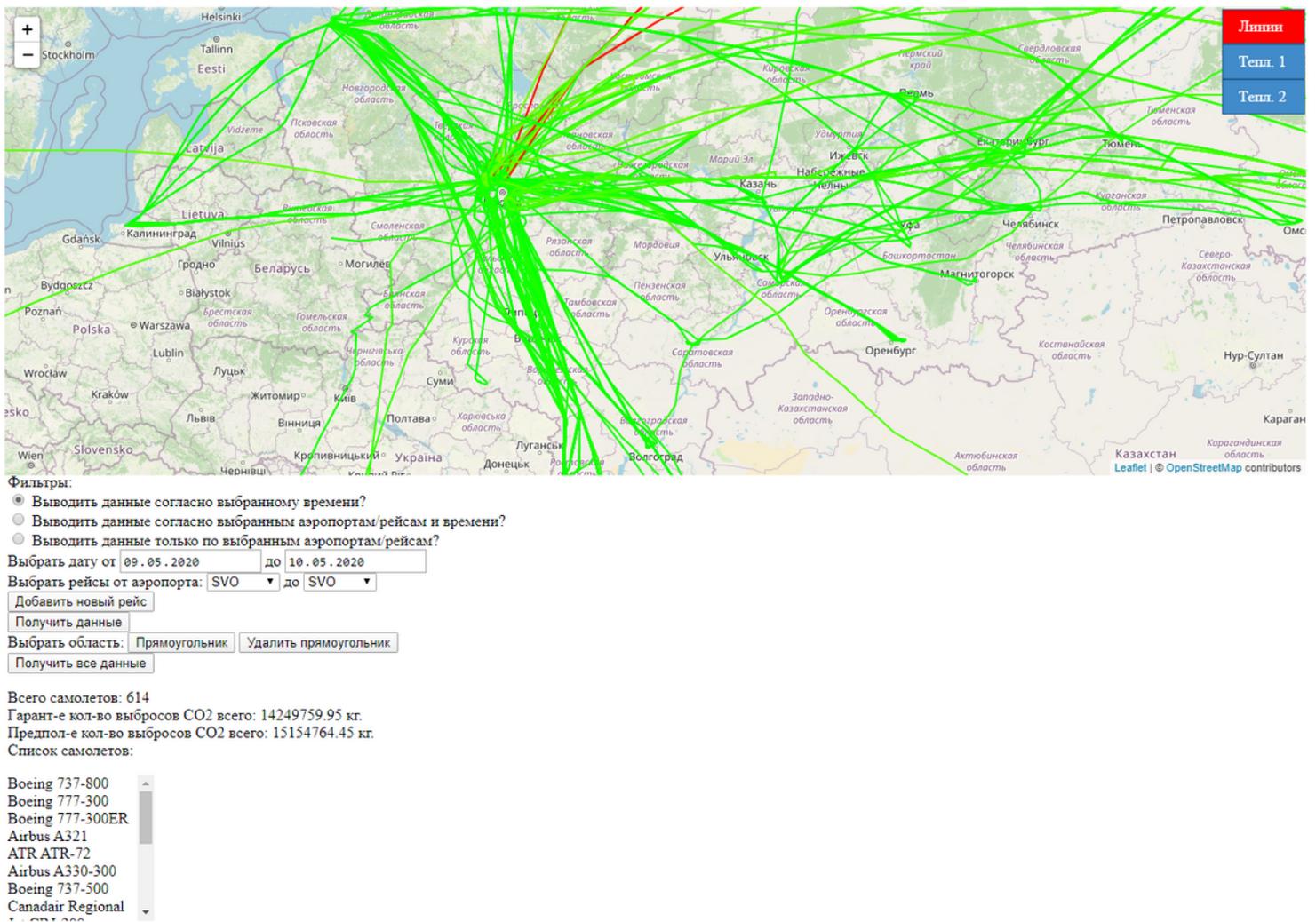


Figure 3

Interactive map interface.

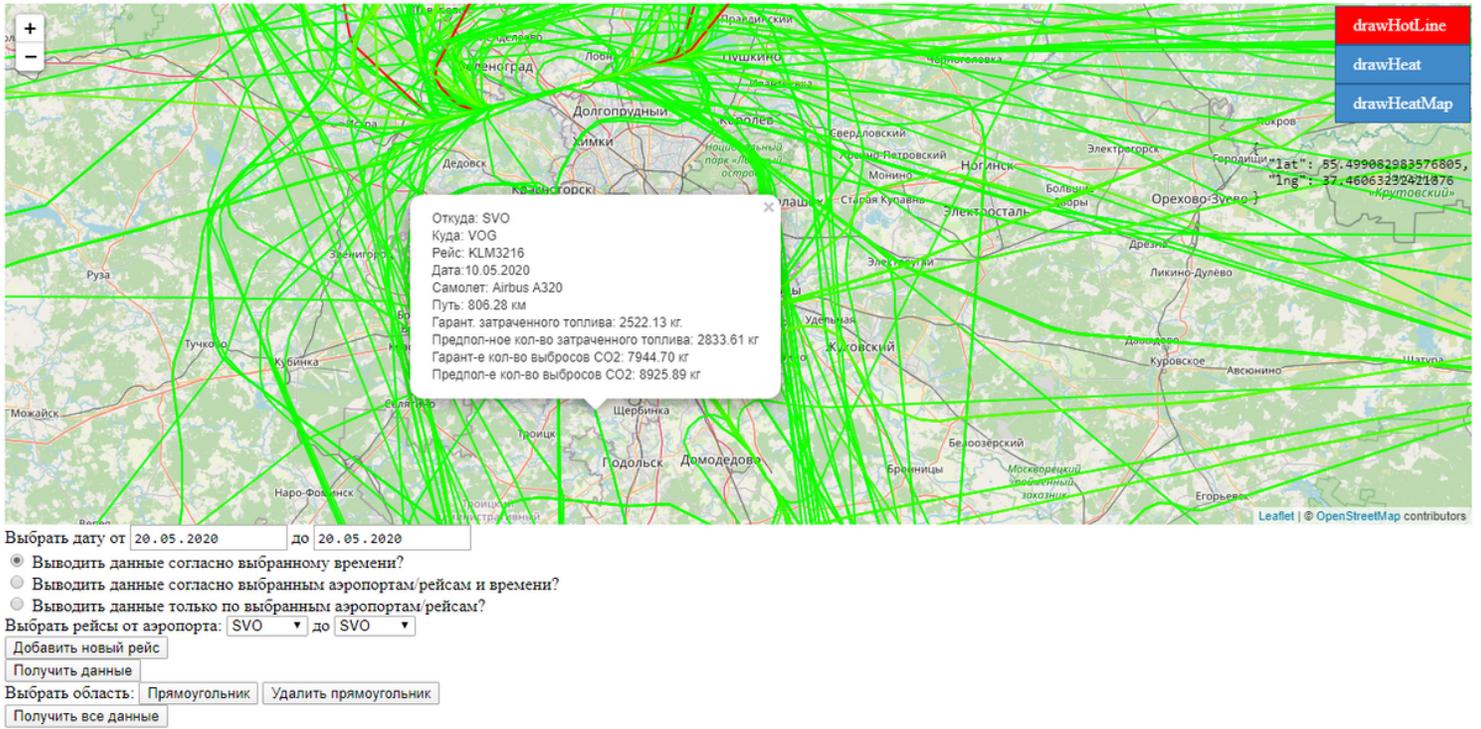


Figure 4

Flight data output.

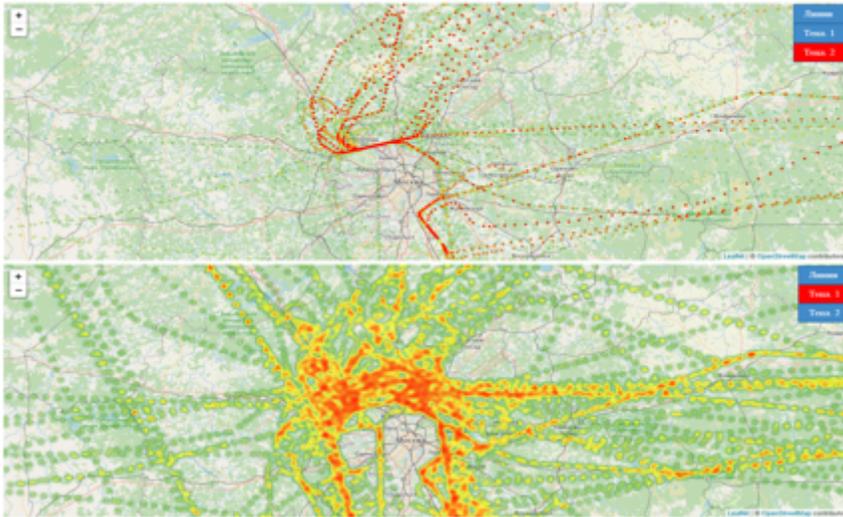
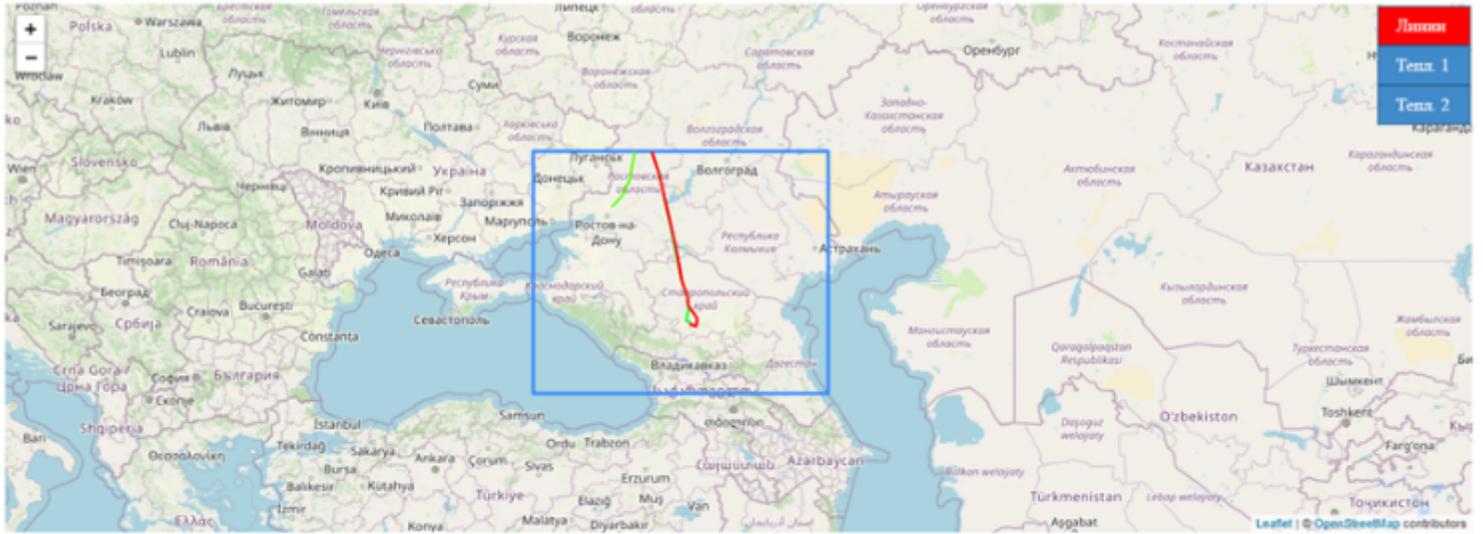


Figure 5

Output of pollution data by flight coordinates from the top, output of a hot map by the number of emissions from below.



Фильтры:

- Выводить данные согласно выбранному времени?
- Выводить данные согласно выбранным аэропортам/рейсам и времени?
- Выводить данные только по выбранным аэропортам/рейсам?

Выбрать дату от до

Выбрать рейсы от аэропорта: до

Выбрать рейсы от аэропорта: до

Выбрать область:

Всего самолетов: 12
 Гарант-е кол-во выбросов CO2 всего: 86029.96 кг.
 Предпол-е кол-во выбросов CO2 всего: 87130.13 кг.
 Список самолетов:
 Airbus A320
 Sukhoi Superjet 100

Figure 6

The architecture of the designed tool.