Big Data Environmental Monitoring System in Recreational Areas

Alexander Volkov[1], Andrey Kopyrin[1], Natalya Kondratyeva[1] and Sagit Valeev[1]

[1] Sochi State University, Sochi, 354008, Russia
`vss2000@mail.ru`

Abstract. The paper discusses the architecture of the environmental monitoring system of the recreational area. It is assumed that the system uses big data technologies to process heterogeneous information. Information in various formats is obtained using remote sensing of the earth, aerial photography, shooting from unmanned aerial vehicles, and internal monitoring of the state of objects of the recreational zone. The area of pollution and the volume of emissions and, consequently, the concentrations created by them are determined by experts considering the available data. The necessity of considering the dynamics of the ecological state based on the available data generated by weather services and remote sensing data of the earth in various spectral ranges is shown. Also, within the framework of the monitoring system, statistical data and results of modeling the ecological state of the recreational zone are used. Based on the analytics of big data, it is supposed to build a forecast of thermal emissions and the amount of harmful substances in the atmosphere carried by air masses.

Keywords: Environmental Monitoring, Big Data, Multi-level Information Collection System, Databases, Data Aggregation.

1 Introduction

As it is known, the main purpose of the recreational zone is to restore the health of the population and ensure the preservation of the labor capital of any state. Assessment of the pollution of the recreational area, considering the scale of environmental pollution, allows you to control the ecological situation in recreational areas within the region.

Assessment of the state of complex systems is based on a hierarchical collection of models of various details [1]. When building a monitoring system for the ecological state of a recreational zone, models of various details are used, based on the analysis of available information.

Thermal emissions and emissions of harmful substances carried by air masses significantly affect the ecological state of coastal recreational zones. These circumstances affect the level of recreational services provided, as well as the ecological state of the resort area. Predicting the ecological state and reducing harmful emissions of various natures is an extremely urgent scientific and practical task [2, 3].

Determining the area of these zones of increased concentration of harmful substances allows us to qualitatively assess the level of pollution of the zone based on

expert estimates and big data analytics. The area of zones with an elevated level of harmful substances at different times is significantly different. This is due to several reasons, among which the climatic conditions of the considered recreational region and the trajectory of the air masses transporting harmful substances over long distances [4-7].

From this we can conclude that it is necessary to take into account the dynamics of the ecological state, taking into account the available data generated by weather services and data from remote sensing of the earth in various spectral ranges.

To assess the permissible concentration of harmful emissions, stationary systems for measuring air quality and monitoring based on aerial photography and satellite images are used.

The area of pollution and the volumetric consumption of emissions and, consequently, the concentrations created by them are determined by experts considering the available data. In total, thermal emissions into the environment and emissions of harmful substances from megacities lead to a significant deterioration of the ecological state of the resort area, which is quite far from the places of emission.

To analyze and make an effective forecast, the problem of integrating systems for collecting and processing a large amount of data is considered.

Fig. 1 shows a map of emissions of harmful substances by megacities in the resort areas of Spain in the winter, shows areas of increased concentration of harmful substances on the coast Z_1 and Z_2. Solid arrows show the vectors of movement of air masses in this period.

The determination of the area of these zones allows a qualitative assessment of the level of pollution based on expert estimates.
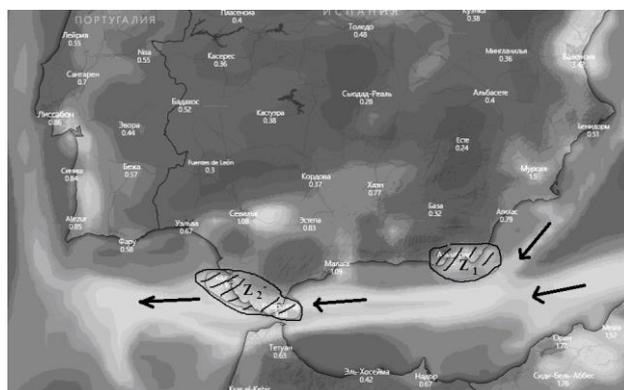

Fig. 1. Map of the distribution of harmful emissions in recreational areas of the coast of Spain in winter.

As follows from Fig. 2, the area of zones with an increased level of harmful substances in the high season is much larger than in the low season. This is due to several reasons, among which the climatic conditions of the considered recreational region and the increase in demand for services in certain periods of the year. From this we can conclude that it is necessary to take into account the dynamics of the ecological state and possible

emissions of harmful substances associated with heating and cooling air in the premises, taking into account the zonal climatic conditions of the recreation area.

The effectiveness of solving the problem of improving energy efficiency, and thereby reducing harmful emissions during heating and cooling can be achieved using modern technologies and methods. Among them are the reconstruction of the existing construction fund, the district heating network and the introduction of new generation cooling systems [8-13].
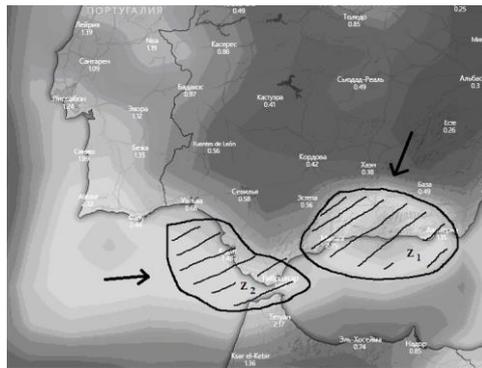

Fig. 2. Map of the distribution of harmful emissions in recreational areas of the coast of Spain in the summer.

It should be noted that a feature of the recreational zones of the seacoasts is the presence in the general fund of recreational infrastructure facilities of many small private hotels with limited opportunities for the application of innovative energy-saving technologies. Another feature of the recreational areas is the specificity of the transport infrastructure, which is overloaded during peak seasons, which leads to traffic jams and increased emissions of harmful substances.

The concentration of emissions significantly depends on the wind rose, as well as on the time of day and time of year. Thus, it should be noted that in the summer season additional active measures are required to control harmful emissions by vehicles.

Thus, the task of creating an information system for monitoring the level of pollution of a recreational zone based on the use of multilevel data analysis is urgent. A feature of the proposed solution is the collection and storage of data from all available information sources and their integration for a step-by-step analysis of pollution sources, as well as taking into account the dynamics of changes in the level of pollution, taking into account the movement of air masses transporting harmful substances from remote industrial regions [14-17].

The principles of designing an information system for environmental monitoring of the recreational zone, which uses data from remote sensing of the earth, aerial photography, surveys from unmanned aerial vehicles, internal monitoring of the state of the objects of the recreational zone, as well as statistical data, were discussed in [18].

Next, we will consider the development of this idea based on a three-level hierarchical information system for collecting heterogeneous information about the ecological state of the recreational zone and algorithms for filtering noise using neural network technologies.

2　　The Task of Analyzing the Amount of Harmful Emissions

Solving the problem of analyzing the dynamics of changes in air pollution in recreational areas, taking into account seasonal temperature fluctuations and the amount of emissions of harmful substances from vehicles, taking into account the influence of the direction of movement of air masses, depending on the season, carrying harmful substances from large cities and industrial enterprises, allows to determine the harm applied to the recreational area.

It is assumed that data are collected both at the regional level and at the level of individual buildings, which will make it possible to determine the level of pollution and their main causes at various levels of the hierarchy, which, in total, can provide an opportunity to control the situation [19-20].

As you know, processing the results of aerial photography in the infrared spectrum allows you to localize sources of heat loss, including in heating networks. The use of small-sized aircraft makes it possible to collect statistical data taking into account the time of day with higher accuracy and at lower material costs.

The problem of determining the area of pollution can be solved by determining the sum of areas of pollution sources. In this case, numerical integration is not assumed. The search can be reduced to the analysis of images with the determination of the total area of contamination by the images themselves.

The total level of pollution taking into account the emissions of heat and harmful substances can be estimated using the following expressions:

$$S_H = \sum_m \oint_H x \, dy \approx S_H^1 + \ldots + S_H^m, \ S_H \leq S_H^S$$

$$S_C = \sum_n \oint_C x \, dy \approx S_C^1 + \ldots + S_C^n, \ S_C \leq S_C^S$$

$$S_O = \sum_k \oint_O x \, dy \approx S_O^1 + \ldots + S_O^k, \ S_O \leq S_O^S \tag{1}$$

$$R = k \cdot S_H + l \cdot S_C + m \cdot S_O$$

where R is the overall level of pollution; S_H is emissions of harmful substances associated with heating; S_C is emissions of harmful substances associated with cooling; S_O is emissions associated with traffic flows; k, l, m are weighting factors, H– heating, C – cooling, O – other loses.

Threshold values of permissible levels of contamination for (1) are determined on the basis of regulatory documents of various departments and organizations.

The choice of the values of the weight coefficients k, l, m depends on many factors determined by expert assessments, various standards, etc.

The volume of emissions is characterized by the area and concentration of harmful substances. It is assumed that the data collection performed by stationary measurement systems allows us to obtain an average value taking into account the season.

Fig. 3 presents a map of emissions of harmful substances in high season. To assess the permissible concentration of harmful emissions, stationary air quality measuring systems are used.
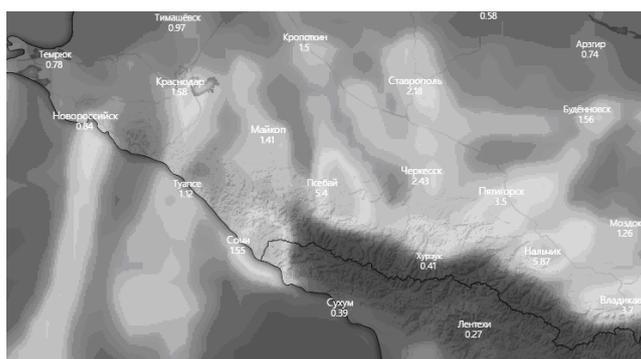

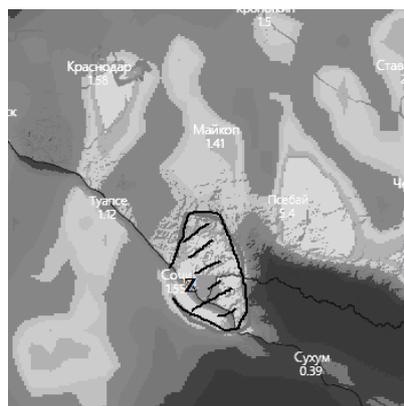**Fig. 3.** Emission distribution map.

Fig. 4. High season emissions.

Area of pollution Z as well as volumetric emissions consumption and, therefore, the concentrations created by them are determined by experts considering the available data (Fig. 4). In aggregate, thermal emissions into the environment and emissions of harmful substances by transport systems lead to a significant deterioration of the ecological state of the resort area. To analyze and make an effective forecast, it is necessary to actively use systems for collecting and processing a large amount of data.

To solve this problem, it is proposed to use a multi-level system for collecting information on the environmental characteristics of the recreational area [4, 5]. The data collection and processing system integrates data arrays obtained by processing infor-

mation from satellite images, distributed monitoring data on the thermal state of infrastructure facilities, as well as the concentration of harmful substances in the resort area at different times of the day and time of the year.

Considering the wind direction makes it possible to improve the quality of the forecast, since it considers the level of pollution, the change in which is caused by air masses carrying harmful substances.

3   The Architecture of the Environmental Monitoring System of the Recreational Area

The data aggregation system includes data processing subsystems obtained on the basis of image analysis and weather observation data. Thus, the problem under consideration is primarily associated with the organization of the collection of information on the ecological state of resort facilities. The complexity of solving the problem under consideration is due to the need to collect process and store heterogeneous information, as well as the need to solve the problem associated with the development of a forecast of the ecological state of the region.

To solve the problems of processing heterogeneous information of large volume, certain hopes are laid on their solution using big data technologies. This is due to the successful practical development of new data transfer technologies, the implementation of data management systems based on NoSQL technologies, as well as the successful implementation of distributed data processing based on cloud computing.

In the framework of the use of big data technologies, the application of algorithms for determining energy losses and pollution levels based on the results of image analysis and statistical results of environmental modeling is promising. In this regard, the development of a system for collecting data on energy consumption can ensure the efficiency of the procedure for developing a plan for the generation of thermal energy for heating and cooling, and optimal plans for loading transport arteries in the region, taking into account the wind rose and climatic conditions.

When collecting data by aircraft flying at different heights, the level of detail of the images obtained changes, which leads to the need to synchronize the received data in time and coordinates. At the same time, the formats for presenting graphical information and data from various sources can differ significantly.

Fig. 5 presents the generalized architecture of a three-level data collection system for monitoring the energy consumption used in heating, cooling facilities of a recreation area in various climatic conditions and collecting data on air quality.

A feature of the organization of computational processes and data storage in the system under consideration is the need to process data on the thermal state of the spa zone, the thermal state of spa resort facilities and the internal thermal state of objects presented in various formats.

The complexity of solving this problem is due to the need for reliable localization of objects and linking all available information about losses to these objects.

In Fig. 5 at the city level L1 (where O_i, $i = 1 \ldots m$, are areas of different energy losses), based on the analysis of data on excess electricity consumption, heat losses in heating

networks, the concentration of harmful substances at the current time, determined by the state of the city's transport routes, it is possible to determine zones that are unfavorable from the point of view of the ecological condition.
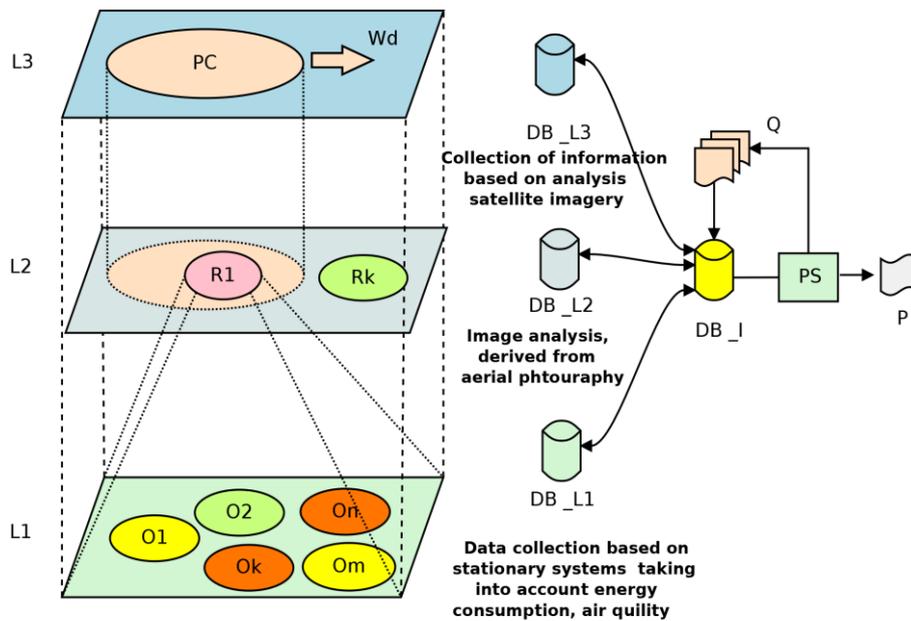

Fig. 5. Generalized scheme of a three-level system for collecting data from monitoring results

Data obtained on the basis of analysis of power consumption, temperature measurements, wind directions, measurements of the concentration of harmful substances with reference to a city map are stored in the DB_L1 database.

To obtain information on the state of the characteristics of the ecological state of the region, statistical information is collected from urban environmental services, as well as information obtained on the basis of aerial photography and images in various spectral ranges obtained using meteorological satellites. The analysis results at this level make it possible to determine the level of pollution in various areas of the recreational zone (R_j, j=1…k). Data at this level is aggregated in the DB_L2 database.

Sources of pollution, which can be attributed to industrial enterprises, transport systems of megalopolises, to a large extent currently affect the ecological state of recreational areas. This is caused by the transport of harmful substances by air masses over long distances (in Fig. 5 PC is a pollution cloud, Wd is a wind direction). Information on the direction of air masses is presented at the L3 level, data on the speed and direction of air masses is in the database BD_L3. The data sources are L1 and L2 weather stations. The results of satellite imagery analysis can also be used. Data considering the heterogeneity of formats, the time of their receipt are placed in the DB_I database. These data are used to build the PS environmental forecasting system. When a request

for forecasting Q arrives, the forecasting system, using the integrated data on the current state of the system L1, L2, L3, issues forecast P.

Several states of the ecological system should be distinguished: normal, harmful to health, and critical. Depending on the results of the analysis, the consumers of this information may be residents of the recreational zone or the relevant state and municipal services.

As noted earlier, data integration means the use of all available data on the ecological state of objects in the recreational zone and the zone itself, as well as solving the problem of linking them to the objects under consideration and to the zone itself. Images obtained using various data collection tools have different resolutions and details. They were obtained at different times and can display different spectral components of the object state. Obtaining a general picture requires solving problems of semantic processing and using various algorithms for image analysis.

Based on the data obtained, it is proposed to construct a textbook for training neural networks, with the help of which it is supposed to obtain a forecast.

Determining the area of objects under study using software systems for analyzing space images involves a large proportion of manual processing using graphic tools built into these systems. When determining the level of pollution on images with a high resolution, it is necessary to solve the problem of assessing the total area of pollution zones. It should be noted that the quality of images depends on many factors: illumination, cloudiness, etc.

The data aggregation system includes subsystems for image processing, determining the level of thermal state of objects, and localization of objects. Aggregated information is placed in the database DB_I.

In the case of processing a video series, the problem of image recognition based on frame extraction arises.

Fig. 6 shows an image analysis algorithm that allows you to determine the quality of satellite images depending on the cloud cover.

The proposed algorithm consists of frame extracting, pre-processing stage, training stage and testing stage. In Fig. 6, TSN segment is a segment of the temporal segment network (TSN), a novel framework for video-based action recognition. It refers to the idea of long-range temporal structure modeling.

It includes a sparse temporal sampling and video-level supervision. Data augmentation is a common technique to improve results and avoid overfitting during model training. To solve the problem of image analysis, a convolutional neural network is used (CNN) [18, 19].

Thus, the problem under consideration is primarily associated with the organization of the collection of information on the thermal state of resort facilities and the state of air quality.
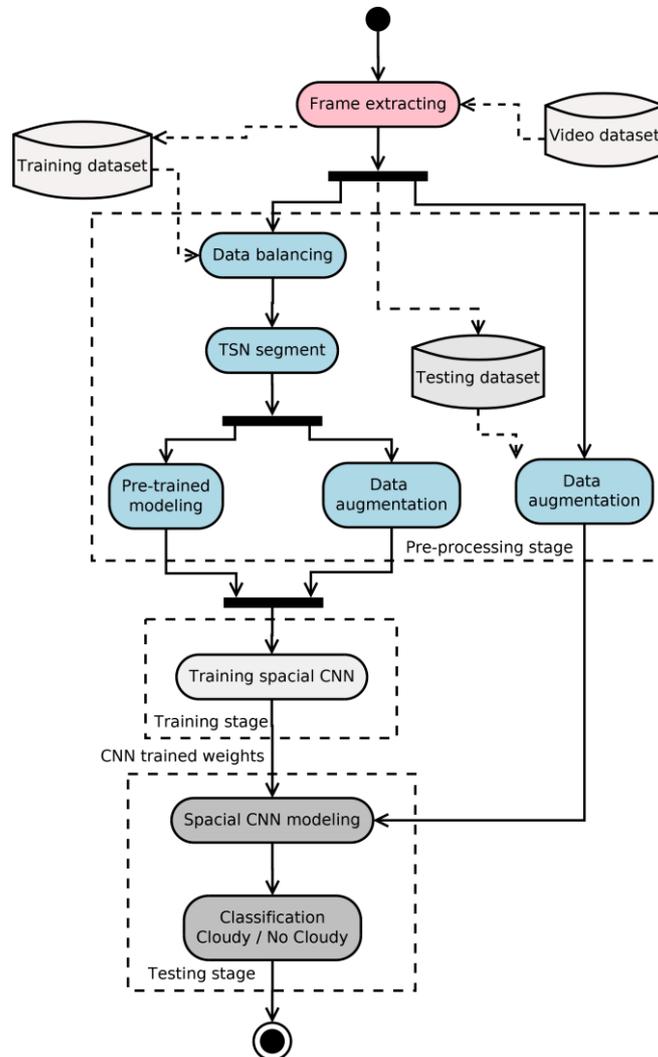
Fig. 6. Algorithm of image classification (Cloudy/No Cloudy)

The complexity of the solution of the problem under consideration is due to the need to store heterogeneous information and the need to solve the problem of localizing resort facilities, determining heat losses in different climatic conditions, taking into account the peculiarities of energy use for various needs.

4 Conclusion

The problem of organizing the collection of information on the ecological state of the resort area is considered. The complexity of the solution is due to the need to organize

the storage of heterogeneous information and solve the problem of localization of resort facilities, determine heat losses in various climatic conditions, taking into account the peculiarities of the use of energy resources, as well as the problem of analyzing the state of air quality in the dynamics of the quality of the environment. A multilevel data collection system using data from remote sensing of the Earth, aerial photography and surveying from unmanned aerial vehicles, as well as data from internal monitoring of both the thermal state of objects in the recreation area and air quality are discussed.

It is assumed that the proposed solutions may be in demand when assessing the harm caused to the recreational area by industrial emissions and insufficient attention to energy-saving technologies for heating and cooling. The results of the analysis can serve as a factual base when searching for solutions at various levels of decision-making, including the regional level.

References

1. Thalheim, B.: Models for Communication, Understanding, Search, and Analysis. In: Proceedings of the XXI International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2019), pp. 3-18 (2019)
2. Overall heat transfer loss from buildings - transmission, ventilation and infiltration, https://www.engineeringtoolbox.com/heat-loss-buildings-d_113.html
3. Theodore, L., Behan, K.: Introduction to Optimization for Chemical and Environmental Engineers, 1st ed. CRC Press (2018)
4. Top six places for energy losses in commercial buildings. https://www.ecdonline.com.au/content/test-measurement/article/top-six-places-for-energy-losses-in-commercial-buildings-259097222
5. Building Envelope: How to Avoid Energy Loss. https://www.facilitiesnet.com/energyefficiency/article/Building-Envelope-How-to-Avoid-Energy-Loss--9428
6. Aerial Infrared Inspection. http://nationwidedrones.co.uk/drone-infrared-inspection
7. Making Your Facilities A Safer Place. http://preciseir.com/
8. Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of massive datasets. Cambridge University Press (2014)
9. Bohlouli, M., Schulz, F., Angelis, L., Pahor, D., Brandic, I., Atlan, D., Tate, R.: Towards an integrated platform for Big Data analysis. In: Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives. Springer, Berlin, pp. 47–56 (2013)
10. Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: DATA exchange: semantics and query answering, in Proc. of the 9th Int. Conf. on Database Theory (ICDT 2003), pp. 207–224 (2003)
11. Jacobs, B.: Categorical Logic and Type Theory. In: Studies in Logic and the Foundation of Mathematics. 141. Elsevier, Amsterdam (1999)
12. Jones, M., Schildhauer, M., Reichman, O., Bowers, S.: The new bioinformatics: integrating ecological data from the gene to the biosphere. In: Annu. Rev. Ecol. Evol. Syst. 37(1), pp. 519–544 (2006)
13. Labrinidis, A., Jagadish, H.V.: Challenges and opportunities with Big Data. In: Proc. VLDB 5(12), pp. 2032–2033 (2012)
14. Lenzerini, M.: Data integration: a theoretical perspective. In: Proc. of the 21st ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2002), pp. 233–246 (2002)

15. Lenzerini, M., Majki´c, Z.: First release of the system prototype for query management. Semantic webs and agents in integrated economies, D3.3, IST-2001-34825 (2003)
16. Lenzerini, M., Majki´c, Z.: General framework for query reformulation. Semantic webs and agents in integrated economies, D3.1, IST-2001-34825, February (2003)
17. Sazontev, V.: Methods for Big Data Integration in Distributed Computation Environments. In: Proceedings of XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), pp. 239-244 (2018)
18. Volkov, A.N., Kopyrin, A.S., Kondratyeva, N.V., Valeev, S.S.: Ecological Monitoring Information System in Recreational Zones, Scientific Journal Engineering System and Constructions, 1(38), pp.20-24 (2020) (in Russian)
19. Kondratyeva, N.V., Valeev, S.S.: Simulation of the life cycle of a complex technical object within the concept of Big Data. In: CEUR Proceedings of 3rd Russian Conference Mathematical Modeling and Information Technologies, pp. 216-223 (2016)
20. Volkov, A., Kopyrin, A., Kondratyeva, N., Valeev, S.: Multilevel Data Acquisition System of Energy Losses in Recreation Areas. In: 2019 Twelfth International Conference "Management of large-scale system development" (MLSD'2019), Moscow, Russia, pp. 1-4 (2019)