Application of Bayesian approach and probability theory for calculating territorial risk and zoning on urban area

Y A Tunakova¹, A R Shagidullin², V S Valiev² and A R Galimova¹

¹ Kazan National Research Technical University named after A.N. Tupolev (KNRTU–KAI), Kazan, 420111, Russia
² Institute for Environmental Problems and Subsoil Use of the Academy of Sciences of Tatarstan, Kazan, 420087, Russia

E-mail: juliaprof@mail.ru

Abstract. The article describes a new algorithm for assessing territorial risk. The risk in the algorithm is calculated as the ratio of the prior and posterior probabilities of an unfavorable (risk) event. An event is an excess of the threshold content of toxicants in a certain area of space, taking into account how often this event occurred in all other areas of space. This can be used for zoning an urbanized area. We used a probabilistic safety assessment (PSA) apparatus based on hazard modelling and a scenario-based approach. Such an apparatus does not analyse certain integral characteristics of the process, but all possible states in a given time section at the same time, forming a probabilistic model. The probabilistic approach ensures the accuracy and realism of assessments of the state of the environment of an urbanized area that are unattainable for deterministic methods and, in addition, allows you to take into account the cumulative effect of multiple sources of uncertainty. Using the probabilistic method and assessing the level of risks allows you to broadcast complex scientific information in an accessible format to people making environmental decisions. Risk is not an absolute but a relative category. It is heterogeneous in both time and space. Therefore, it can be used when zoning the environment, especially since the parameters that are used to calculate risk very often form the same data arrays necessary for classifying and zoning spatial non-uniformity of the system in general. After analyzing various approaches, we came to the conclusion that for a given time period the probability of an adverse event will be calculated more accurately if we calculate not using the classical formula (the number of cases per time interval), but in the form of the ratio of the posterior and prior probabilities. This takes into account the risk in a certain area of space and correlates with the frequency of the same event in all other areas. To assess the likelihood of polymetallic pollution of environmental objects in an urbanized area with its subsequent zoning, we considered the variability of the metal content in environmental objects. Based on the probabilities that metal concentrations exceeded their threshold values, the territorial relative risk was calculated and zoning of the urbanized territory was carried out. Testing of the considered approaches was carried out in the territory of Kazan on the basis of an array of data of experimental measurements on the content of metals in environmental objects. For sampling, 170 points were selected across the territory of the city of Kazan, which are under the influence of stationary and mobile sources of the influx of metals into the urbanized territory. The probability distribution of exceeding the threshold level of metal content in environmental objects in the territory of Kazan is shown cartographically.
1. Introduction
A huge number of multidirectional and different-sized factors acting in an urbanized area, the heterogeneity of their distribution in space, non-linear relationships between them, creates significant difficulties for assessing environmental risk with territorial differentiation. The assessment of environmental risk, according to [1], determines the likelihood that adverse environmental effects can occur as a result of exposure to one or more sources. Currently, among the methods for assessing environmental risks, two groups of methods are most often used, called engineering and probabilistic.

Engineering methods are based on a logical-probabilistic consideration of the scenario of the development of events. An example of the implementation of these methods is the construction of an "event tree" and "fault tree" based on a graphical representation of the probability of each unwanted event. These methods can be used for both quantitative and qualitative assessment of environmental risk. But for the development of control actions, this method is not very applicable due to the high labor intensity of assessing the probability of all possible scenarios for the development of events when exposed to risk factors in an urbanized area.

Probabilistic methods are based on the application of algorithms of probability theory to datasets of measurements of certain quantities or events (statistical series). They allow both to calculate the probability of a negative event occurring, and the likelihood of avoiding this negative impact. The advantage of using statistical data series is the ease of interpretation of the resulting environmental risk assessments. Accordingly, the main limitation of the method is associated with the lack of statistical data. [2]. In simpler terms, the assessment of the environmental risk for the measured quantities (for example, the levels of toxic substances in the atmosphere or river) can be calculated as the ratio of the actual level of the quantity (chemical substance, exposure) to the permissible normal value (most often MPC, toxicity). [3]. Based on this principle, international standards have been developed in the field of risk analysis and assessment of the likelihood of adverse events. Statistical characteristics (mathematical expectation, median, quantile, etc.) of probability distributions of the corresponding random variables are used as quantitative estimates of the gradations of the level of environmental risk. The disadvantage of this approach is that the environmental risk assessment (ERA) is calculated either for only one or only for a small number of substances (toxicants) and a small number of objects (rivers, soil areas, etc.) that they affect [4].

Within the framework of the probabilistic method, such approaches as first- and second-order Monte Carlo Simulation, Sensitivity Analysis, Interval Analysis, Qualitative Modeling, Aikake Information Criteria, Probability Bounds Analysis, Information Gap Theory and Bayesian Belief Networks, hierarchical Bayesian techniques, etc. Moreover, it is the Bayesian approach that has significant efficiency for making decisions on territorial management.

2. Probabilistic safety assessment
Taking into account the considered advantages and limitations of the methods for assessing environmental risks, an approach was developed for assessing the territorial relative environmental risk, which allows using the results of the an assessment for making recommendations for the management of an urbanized area. For this we used probabilistic safety assessment (PSA) methods. In contrast to methods that do not use probability, this approach more accurately calculates assessments of the state of ecological objects in the city. In addition, estimates are obtained close to the real situation to a large extent. The probabilistic approach also allows one to take into account the combined influence of many unknown causes due to which a hazard arises.

In general terms, risk is always represented as a function of the probability of an event on the consequences of the implementation of this event. In the simplest case, this is the product of the frequency of an event by the frequency of an adverse manifestation from this event. We analyzed different approaches, after which we came to the conclusion that for a given time period the probability of an adverse event will be calculated more accurately if we calculate not using the classical formula (the number of cases per time interval), but in the form of the ratio of the posterior and prior probabilities. This takes into account the risk in a certain area of space and correlates with
the frequency of the same event in all other areas. Such models, which are to the modeled object in relation to probabilistic similarity, are called stochastic. Unlike deterministic models, they operate with random variables, considering the output result depending on the level of uncertainty. For stochastic models, the probabilities of these results are calculated, respectively, we can assert about one or another degree of determinism of the observed process. Such models are very useful when you need to investigate how the mathematical expectation of random variables adequately represents the real situation. Does the mathematical expectation reach a given limit value? If so, how much is this related to random fluctuations in the measured values? It is in such a situation that we propose to use probabilistic estimation models, in which the experimental dataset can be denoted as D, and the event that the measured parameter has reached a limiting value, as S. Then the conditional probability of the event "For a dataset with a value has reached the limiting value" can be denoted as p(S|D).

Let us express the total probability of event S:

\[ P(S) = \sum_{i=1}^{n} p(D_i)p(S|D_i) \]

Events \( D_i \), (i=1, 2, 3...n) form a complete group:

\[ \sum_{i=1}^{n} p(D_i) = 1 \]

In this case, it is obvious that the event S occurs in any case. In this case, it will happen in one of the \( D_i \). Then the probability is determined by the formula:

\[ P(D_i|S) = \frac{p(D_i)p(S|D_i)}{p(S)} \]

Probability \( p(S|D) \) term the posterior probability, and \( p(S) \) — prior probability. Here, the denominator of the expression contains the integral for continuous, and the sum for discrete random variables, that is, the integral of the numerator over all possible parameter values. This is necessary to normalize the posterior probability.

In order for us to be able to use the probabilistic model in real life, first of all, it is necessary to determine what event S is. In our case, this is the fact that any parameter exceeds its specified maximum value (for example, values corresponding to 0.5- quantiles or predefined regional standards), as well as events \( D_i \) — which indicate that the excess has occurred in the i-th zone, previously formed using the classification of influencing factors. Then p(D_i)=n_i/N, where N — sum of all n \( (N=n_1+n_2+...n_i) \), and p(S|D_i)=1.

As a result, we use the probability of exceeding the maximum permissible value for any measured parameter as an integral (generalized) indicator. Then we can calculate the probabilities of the parameters exceeding their threshold values in each separately taken area.

The probabilities of events opposite to events \( S_1, S_2, S_3 \), since we are considering cumulatively independent events \( S_1, S_2, S_3 \), then respectively equal \( q_{S_1}=1-p(S_1) \), \( q_{S_2}=1-p(S_2) \), \( q_{S_3}=1-p(S_3) \), since the overall probability that the event S, will or will not occur is 1, that is \( p_1+q_1=1 \). The required probability \( P(S) = 1 - q_1 \times q_2 \times q_3 \), according to the theorem of multiplication of probabilities for independent events and the theorem of addition of probabilities for joint events. In other words, we first calculate the overall probability that no excess will occur \( Q=(1-p_1) \times(1-p_2) \times(1-p_3) \), and then we calculate the probability that this will not happen, that is, at least one excess will be noted: \( 1 - q_1 \times q_2 \times q_3 \). Thus, based on the probabilities of exceeding the concentration of metals in environmental objects of their threshold values, we calculate the territorial relative environmental risk and carry out zoning of the urbanized area.
Our proposed approach for assessing the territorial relative ecological risk within the urban ecosystem of a large city (Urbo-Scale Risk Assessment) should be primarily used for industrial areas, where, on the one hand, there are many sources of pollution, and on the other, there are many objects on which they affect. The approach is based on a relative risk model (RRM). The main difference of the proposed approach is the construction of a generalized city map with an assessment of the territorial ecological risk of polymetallic pollution. By means of mapping, one can characterize the integral degree of pollution, highlight zones of increased pollution and epicenters of pollution, and also assess the contribution of the main sources of pollution [5]. The parameters that are used to assess environmental risk very often form the same data sets necessary for the classification and zoning of the spatial unevenness of the urban ecosystem in general.

3. Results of assessing the territorial relative environmental risk

The aerogenic input of pollutants into the urbanized area is dominant and characterizes the intensity of anthropogenic impact [6]. The calculation of the probability of exceeding the threshold values of concentrations was carried out in relation to the content of metals on the earth's surface. Metals are among the priority pollutants for controlling airborne pollutants with emissions from stationary and mobile sources of pollution.

Metals without transformation migrate from atmospheric air to the underlying surface - soil and snow covers, accumulating in them. In this regard, the initial data array was formed on the basis of our chemical-analytical determinations of the content of metals-components of emissions in the territory of Kazan (Cd, Pb, Co, Cu, Ni, Zn, Cr, Mn, Fe) in snow (μg / l) and soil cover (mg / kg) (indirect characteristics of the level of pollution of the surface layer of atmospheric air), with statistical processing of the results. For sampling, 170 points were selected on the territory of Kazan, which are under the influence of stationary and mobile sources of metal emissions (figure 1-2).

![Figure 1](image1.png) **Figure 1.** Distribution of probabilities of exceeding the threshold level of metal content in the soil cover in the city of Kazan.

![Figure 2](image2.png) **Figure 2.** Distribution of probabilities of exceeding the threshold level of metal content in snow cover in the city of Kazan.

Analysis of the maps shows that the epicenters of pollution are mainly tied to the location of stationary emission sources operating in the central part of the city, while the risk of snow cover pollution is expected to be a more dynamic indicator than the risk of soil pollution, which makes it possible to identify areas of pollution with mobile sources. Differentiated assessment of the probability of contamination of snow and soil cover allows for a clearer interpretation of the distribution of risk zones in the territory of the urban ecosystem. Superimposing high probabilities calculated for both snow cover and soil allows the dirtiest areas to be identified. Areas with a high probability of contamination of only snow cover are characterized by pollution mainly from mobile sources, and areas with a high probability of soil contamination are characterized by long-term exposure to
stationary sources. Relatively clean areas, which include the eastern and southeastern districts of the city, are characterized by low values of the probabilities of contamination of both soil and snow cover.

4. Conclusion
Thus, we have proposed an algorithm for assessing the territorial relative environmental risk as a ratio of a posteriori and a priori probabilities, taking into account the probability of a risk event - exceeding the threshold metal content for the considered list of metals (Cd, Pb, Co, Cu, Ni, Zn, Cr, Mn, Fe) in a specific area, regarding information about how often this event occurred in all areas of space. The results of approbation of the developed algorithm for the zoning of an urbanized territory on the example of the city of Kazan are obtained. The cartographically presented distributions of the probabilities of metals exceeding their threshold values make it possible to differentiate zones of pollution by stationary or mobile sources on the territory of the urban ecosystem and to highlight areas with a high degree of polymetallic load. The carried out complex zoning of the territory of the city of Kazan according to the degree of polymetallic pollution by metals can be used for the ecological justification of schemes for the development and placement of production facilities, master plans of cities, etc. It can be concluded that in areas of a high level of territorial relative environmental risk, landscaping of the territory is especially significant and the placement of residential buildings should not be planned. In areas with a significant contribution from mobile sources of emissions, it is necessary to develop traffic control schemes to reduce traffic loads. For zones with a predominance of the influence of stationary sources of emissions, the emergence of new sources of pollution should be excluded.

Funding
The research was carried out with the financial support of the Ministry of education and science of the Russian Federation in the framework of fulfilling obligations under Agreement number 075-03-2020-051/3 dated 09.06.2020 (topic number fzsu-2020-0021).

References
[1] Guidelines for Ecological Risk Assessment 1998 United States Environmental Protection Agency
[2] Ecological Risk Assessment Guidance. Federal Contaminated Sites Action Plan (FCSAP) Azimuth Consulting Group Inc., Vancouver 2012215
[3] Verdonck F 2002 Probabilistic ecological risk assessment framework for chemical substances Proceedings International Conference on Integrated Assessment and Decision Support 1 144–49
[4] O’Brien G and Wepener V 2012 Regional-scale risk assessment methodology using the relative risk model (RRM) for surface freshwater aquatic ecosystems in South Africa Water SA 38(2) 153–65
[5] Tunakova Y 2000 Environmental and hydrometeorological problems of large cities and industrial zones International Conf. Proc. 51-52 (in Russian)
[6] Tunakova Y A, Novikova S V, Faizullin R I and Valiev V S 2018 Approximation of the dependency of trace elements concentrations in internal media upon their contents in environment objects BioNanoSci 8 288–95