Method for analyzing spatial organization of information for predicting anthropogenic impact on environment

V V Gorin¹, N V Khvatysh¹, T A Sokolova¹, S G Komarova²

¹State University of Land Use Planning, 15, Kazakova street, Moscow, 105064, Russia
²Russian University of Chemical Technology named after D.I. Mendeleev, Miusskaya square, Moscow, 125047, Russia

E-mail: KhvatishNV@guz.ru

Abstract. In the emerging ecological crisis, there are no analysis methods enabling to assess the degree of environmental hazard from a unified position. Moreover, when new characteristics are obtained, the idea of the phenomenon under study can radically change. The article considers the method of informological analysis based on the study of the structure of the spatial organization of information in urbanized territories. In the course of the study, it was shown that the information field formed by natural objects and the technosphere interact with each other in such a way that they form their internally organized information space. The information field inside is estimated according to the results of production control over technological parameters of an enterprise. According to its state inside the object, zones of information intensity of the object itself are identified, and external feature fields contribute to describing the impact of an individual enterprise on the environment of territories limited by the size of each feature information field. The latter is the initial element of the mechanism for transforming information of an industrial facility situated at a larger space, which may include interconnected objects of the industry structure, cover zones of characteristic environmental changes not only within the location of an industrial territorial unit but also outside it. In these zones, fields of information interaction being the factors of environmental hazard although often not taken into account by traditional methods of analysis can generate. A probabilistic-theoretical model of the message source for calculating the entropy state of space is proposed, which allows describing the characteristic features of the influence of an object on the environment of the territories. A specific example shows its fundamental applicability to studying information aspects of phenomena of various nature, using entropy to identify important properties of the process under study with an unknown model. Further development of the method can be used in various ways for determining environmental hazard factors, which include spaces that form around objects and have a direct or indirect impact on the environment on a local, regional and global scale.

1. Introduction

Despite the fact that the technosphere functions in accordance with the requirements of the norms and rules of environmental safety, which were laid down during its formation period and implemented in relevant projects, at present none of them fully meets the safety requirements. The current situation in the field of ecology and protection of public health has reached critical values. There are a number of reasons for this. The main one is as follows: mankind does not have sufficient theoretical and practical
experience in solving problems of this scale and complexity since the existing methods for analyzing the processes occurring in nature were created during a period of relative stability of the planet’s ecological state. The results obtained were based on the analysis of the sequential interaction of various characteristics with each other, which gradually accumulated to fill the required and sufficient volume. Such an analysis scheme requires a valuable time, which is quite acceptable during a stable ecological state and extremely unacceptable under conditions of a developing ecological crisis.

The disadvantages of existing analysis methods are the impossibility of presenting the problem as a coherent whole. Thus, it seems possible to unambiguously determine the degree of environmental hazard and health protection, and, therefore, to offer the only optimal solution for it. As practice has demonstrated, an equally significant drawback is that when new characteristics are obtained, understanding the nature of the phenomenon under study can radically change. This leads to the fact that the probability of making erroneous decisions, in which an ecological crisis can become an inevitable reality, is deliberately included in the analysis mechanism. And, finally, the existing analysis methods have a rather narrow specialization due to being created within the framework of the concept of highly specialized scientific disciplines.

By definition, the environment is a set of bodies, forces, natural phenomena, matter, space and any human activity that are in direct contact with living organisms. The concept of environment includes a natural framework, a system of linear spatial relationships ordered by the degree of ecological value, natural opposition to complex anthropogenic expansion in a single territory [1].

Anthropogenic human activities significantly change the natural composition of foreign substances contained in the environment (water, soil, air and living organisms). These changes enable them to enter into chemical and physicochemical interactions with biological objects of the ecosystem. Chemicals accumulating in the environment in quantities unusual for it and changing natural composition act as eco-pollutants. The change may be caused by excessive accumulation of one or many eco-pollutants in the environment.

The elements of the environment interact with each other by exchanging mass, energy, information. Thus, the environment acts as a geophysical and biological diversity with or without elements of technogenesis. It has n-sensitive elements, each of which can be influenced by m-external and internal natural and anthropogenic factors.

As a result of exposure, the environment can react by n×m responses belonging to different classes according to certain characteristics or properties that differ from each other, which can be considered as a variety of elements states and their relationships.

Consequently, with all the differences in the interpretation of the concept, information always manifests itself in material and energy form represented by various signals. In this case, the source of messages in accordance with its design, purpose and method of physical implementation at a fixed moment in time selects one of the message sets. In a particular case, a measuring device, an information-measuring system, or a person performing the function of an observer can act as a source of information.

Currently, there are no methods of analysis to solve the problems of assessing, predicting the integral characteristics of natural processes and phenomena. This article proposes a method of analysis based on studying the structure of the spatial organization of information in urbanized areas.

2. Technosphere as a source of information about anthropogenic impact on the environment

The existence of information processes in the environment along with the flow of energy and the circulation of matter is recognized by most ecologists. However, what information in ecosystems, geospheres and technosphere is and how the mechanism of its circulation works remains a poorly studied problem.

In the spatio-temporal continuum of technogenesis, there are generators of the material world in the form of three types of geospheres and technosphere. These generators interact with each other. The main driving force is solar energy, which acts as a regulator of the exchange between the elements of the geospheres and the formation and development of the biosphere. The main driving force behind
the evolution of the technosphere is energy obtained using renewable and non-renewable natural resources. At the same time, the abiotic, biotic and anthropogenic components interact at the material, energetic and informational levels.

If the concept of information is associated with natural diversity, then information can serve as a measure of heterogeneity in the distribution of energy or matter in space and time. Information in general form exists abstractly, and all natural phenomena are caused by the interactions of structural and operational information and their mutual transitions (transformations) into each other.

At the same time, the information field formed by natural objects and the technosphere, which interact with each other, have the organization of their information space. This contributes to solving a number of important practical problems in various fields of natural, exact, humanitarian and social sciences using universal informological models of space, information and information field [2].

The structural elements of the technosphere create man-made and environmental hazards arising from environmental pollution with various wastes and energy flows from the work of industrial, service and household enterprises. Zones of anthropogenic impact extend to the regions of the technosphere and adjacent natural zones, to the territory and objects of the economy, transport, urban and residential areas. In some cases, these dangers arise at the interregional and global levels [3].

The spatial combination of production branches and industries is influenced by many factors like provision of mineral resources, fuel and energy, material and labor resources. These factors are closely related to each other and have a certain impact on the location of enterprises and various sectors of the economy. In the process of placing industrial production, various forms of its territorial organization have developed.

Large economic zones are vast territorial entities with characteristic natural and economic conditions for the development of productive forces.

Industrial areas are large territories with relatively homogeneous natural conditions, a characteristic direction of productive forces development, a corresponding existing material and technical base, production and social infrastructure.

Industrial agglomerations are territorial economic entities characterized by a high level of concentration of enterprises in various sectors of the economy, infrastructure facilities and scientific institutions as well as a high population density. The economic prerequisites for the development of industrial agglomeration are a high level of concentration and diversification of production, as well as the possibility of the most efficient use of industrial and social infrastructure systems.

Excessive development and concentration of production beyond certain limits have a negative impact by significantly reducing the economic effect. This is primarily due to environmental issues and social sphere development [4].

An industrial hub is considered as a group of industries compactly located in a small area. Its main features are participation in territorial division of labor, production ties between enterprises, settlement system commonality, social and technical infrastructure.

Considering the above sectoral structures from a systemic perspective, first of all, it is necessary to investigate those aspects and qualities of these objects, which contribute to their ordering and organizing with regards to such basic systemic features as integrity, structure, interdependence on the environment, hierarchy.

Any system in its essence is an information entity while its elements are interconnected through the information flows (in addition to mass and energy). Therefore, information is an objective characteristic of material objects, it provides interaction, dominates at all levels of matter organization and underlies self-regulation and management in nature and society.

Formally, the system structure can be represented as an ordered pair $S = N \subset R$, where $N$ is a set of system elements, and $R$ is a set of relations between these elements.

Let us consider each element of the system as a stochastic signals source, which is in one or another state at every moment with some probability. This state is registered in one way or another (with this or that device) as a message about the internal state of a natural or technical object.
At present, despite the vast experience of studying and creating complex systems for converting information, the classical model of K. Shannon being a universal formula for the amount of information per message remains relevant [5].

\[ I = - \sum_{i=1}^{k} p_i \log p_i \]

where \( k \) is an elementary signal type, \( p_i \) is occurrence probability of the \( k \)th signal.

Having determined the system based on the flow of messages about the internal system state, it is possible to draw a conditional closed boundary, beyond which there are informationally significant elements whose signals are absent. The totality of these elements that influence the system or, conversely, which are affected by it, refer to the external environment. The border between the system and its external environment is very arbitrary because they are always connected by a multitude of interdependent relationships, informational interactions. There is no formal way of defining the boundaries of a system, especially since any system can be an element of another, higher level.

If any particular system is considered as a subsystem of the structure of a higher level, the infinity of space benefits the continuation of this hierarchical process to the level where the physical influence or the possibilities of measuring equipment and technical means of receiving messages extends. If we now perform the inverse operation, specifically, analyze the system by its delayering, then at each level there will repeatedly appear another space (or subspace) of another system. Theoretically, this procedure can be carried out until the moment when the object itself is in the center of the model as a source of anthropogenic impact on the environment. In this case, it is defined as a zone of influence and the environmental hazard factor is assessed using well-known methods of analysis [6].

In informology, such zones refer to spaces with an organized internal state. In this case, the size of the information field depends on the capacity of the industrial facility as well as the size and composition of the main elements and their design. Moreover, the information field inside is estimated according to the results of discrete and continuous production control of technological parameters of the enterprise, processing and analyzing the incoming message flow using a set of methods of multivariate statistical analysis of time series and geostatistics [7]. The information intensity zones of the object are identified based on the state of the information field inside it. External feature fields allow describing the impact of an individual enterprise on the environment of territories limited by the size of each feature information field. The latter is the initial element of the mechanism for transforming information of an industrial facility covering a larger space, which may include interconnected objects of the industry structure [8].

The next step will be the transformation of the information space within new boundaries. This transformation accumulates incoming messages from the original space with an organized internal state, the results of production control from objects additionally included in the external space and the flow of messages coming from the environmental monitoring system. As a result of the transformation, the position of the outer border of the information space can reach several hundred kilometers and cover zones of characteristic environmental changes not only within the location of an industrial territorial entity but also outside it. In these zones, information interaction fields often not taken into account by traditional methods of analysis although being the factors of environmental hazard can generate. Most likely, this is due to the phenomenon of superposition of the transformed information of two or more industrial facilities. Accordingly, the deterioration of the ecological situation is more intense in the territories belonging to these fields. These are so-called abnormal zones. In these zones, unknown diseases in people, animals and plants, abnormal climatic processes can occur inexplicably and the accidents at product pipelines, enterprises, etc. happen more frequent.

Thus, it is possible to expertly assess the state of the fields of information interaction of various production facilities, forecast the development of events for the territories which are located in these fields and whose probability is rather high, using knowledge about the mechanisms of interaction of spaces in which the process of converting information of industrial facilities of various industry structures takes place.
The above theoretical description of the informological analysis method for assessing the consequences of the anthropogenic impact of the industrial sector on the environment can be represented by a quantitative probabilistic model for describing the source of messages in relation to predicting environmental hazard.

3. Quantitative probabilistic-theoretical model of messages source

Let us define the space part containing the signal source as part of the information field supplying messages about the OS state. Let \( X = \{ x_1, ..., x_n \} \) contain \( n \) messages. In this case, the probability distribution \( p(x_i) \) is given on the set \( X \), and

\[
p(x_i) \geq 0, \quad i = 1, 2, ..., m
\]

\[
\sum_{i=1}^{m} p(x_i) = 1
\]

Let \( A \) be a subset of the set \( X \), \( A \subset X \). Then the number

\[
P(A) \triangleq \sum_{x_i \in A} p(x_i)
\]

represents the probability that a random selection from \( X \) in accordance with the distribution \( p(x_i) \) will select a message belonging to the set \( A \) with probability \( P(A) \).

From the perspective of information theory, a source is considered to be completely given if there is a certain probabilistic scheme that enables to calculate the probability of any segment of a message. For example, for a natural object the description is completed in a way enabling to calculate the probability of any combination of parameters characterizing the state of the source at a given point in space at any time.

Therefore, if the source is given, then for any \( n \) and \( i \) and any sequence of messages \( (x^{i1}, ..., x^{in}) \), the probability \( p(x^{i1}, ..., x^{in}) \) of this sequence is determined. The converse is also true. Consequently, all sources that can have a completely different nature given by the same set of probabilities of a sequence of messages are identified from the standpoint of information theory. Moreover, the definition of the source and the definition of the random process that generates the source are the same.

The information field has a number of features that benefit time estimates of the moments of a change in the scenarios of random events development. The concept of entropy closely related to information and introduced by Shannon [5] has become widespread in the interpretation of natural phenomena.

Methods for finding entropy, as a rule, are based on a priori estimates of the events that make up the phenomenon under study [9].

Let \( \{ X, p(x) \} \) be a set of messages obtained as a result of independent observations in the form of either time series or a discrete set of experimentally found measurement points. The result of each observation can be encoded as a sequence of zeros and ones. Then the amount of self-information (a priori uncertainty) stating the occurrence of message element \( x_i \) in the message \( x_i \in X \) is determined by the value \( I(x_i) \)

\[
I(x_i) \triangleq - \log_2 p(x_i), \quad i=1,2,...,L
\]

where \( p(x_i) \) is a priori probability of occurrence of a message element \( x_i \).

The basic properties of a priori information follow from expression (1). Self-information is not negative. The intrinsic information for each message from the set \( \{ X, p(x) \} \) is as follows

\[
I(x_i, y_i) \triangleq - \log_2 p(x_i, y_i)
\]

If \( X, Y \) are independent \( p(x,y) = p(x)p(y) \), then

\[
I(x,y) = - \log_2 p(x) - \log_2 p(y) = I(x) + I(y)
\]

The amount of information is a real function on the set \( \{ X, p(x) \} \) and, therefore, is a random variable with the values \( I(x_1), ..., I(x_L) \).

The mathematical expectation \( H(x) \) of the random variable \( I(x) \), defined on the set \( \{ X, p(x) \} \), determines the following entropy of this set

\[
H(x) \triangleq M I(x) = - \sum_{x_i \in X} p(x_i) \log p(x_i)
\]

Entropy is the average amount of self-information in a message \( X \).
Along with a priori concept of entropy based on probability, it is possible to introduce the concept of statistical entropy or posterior entropy based on the calculation of the relative frequencies of events that correspond to the processed experimental data.

Let us consider a discrete stream of events, \(M(t_1), M(t_2), \ldots, M(t_n)\) that occurred at times \(t_1, t_2, \ldots, t_n\). If the relative frequencies of these events \(W(t_1), W(t_2), \ldots, W(t_n)\) are known, then the information entropy \(S\) can be determined using the following expression:

\[
S = - \sum_{i=1}^{n} W(M_n) \log_2 W(M_n),
\]

To search for information and information entropy that correspond to an ordered set of points, it is necessary to decompose this set into a finite number of subsets, each of which defines one event independent of the others. A set of neighboring points with a common feature can be taken as a separate subset corresponding to its only event independent of others. For plane continuous curves, this feature will be the monotonicity property.

The paper presents the results of processing seismograms using the above-described method of statistical entropy, which shows its fundamental applicability to the study of information aspects of phenomena of various nature [10]. Using this method, the continuous seismogram was transformed into entropy. As a result, the dependence of the relative error of the statistical entropy in time was obtained. The outburst of the relative error of entropy on the curve indicated that the field of information interaction emerged as a factor of environmental hazard, 9 units of time before the onset of the earthquake [11].

This example shows its fundamental applicability to the study of information aspects of phenomena of various nature. This does not require knowledge of the initial model of the studied phenomenon. This approach is based exclusively on empirically obtained discrete sets of points or piecewise continuous curves. Thus, entropy enables to reveal important properties of the studied process with an unknown model: the degree of its so-called overorganization or the degree of its disorder.

This method for determining the statistical entropy in its further development can find a wide range of applications to determining environmental hazard factors. These factors include spaces that are formed around objects having a direct or indirect impact on the environment on a local, regional and global scale (zones of influence). Many facilities of this kind – energy, industrial, transport, etc. – form spatial information fields with an organized internal state. The radius of the information field depends not only on the energy power consumed or generated by the enterprise but also on the main elements of its technological processes and design. The state of the information field corresponds to the state of the information intensity of the object, which will accordingly change the value of the statistical entropy. Analysis of the entropy state of space will allow describing the characteristic features of the impact of an object on the environment of the territories.

4. Conclusion

Thus, the above approach to the consideration of the quality of the environment based on the properties of information and entropy allows solving the problems of complex processing of the measurement results concerning the characteristics of the environment obtained by various methods without resorting to the initial models of the phenomenon under study. In this case, the problem is reduced to the analysis of random processes aimed to obtain the a priori or a posteriori probability of the occurrence of a particular event or phenomenon with the following calculation and analysis of statistical entropy.

References

[1] Federal Law "On Environmental Protection" (as amended on June 25, 2012). dated 10.01.2002 No. 7-FZ.
[2] Sokolov V E (ed) 1992 Ecoinformatics. Theory Practice Methods and Systems (St. Petersburg: Gidrometeoizdat) 519 p
[3] Astakhov A S, Dikolenko E Ya, Kharchenko V A 2018 Environmental safety and efficiency of nature management (Vologda: Infra-Engineering) 323 p
[4] Kalygin V N, Bondar V A, Dedeyan R Ya 2018 *Life safety. Industrial and environmental safety in technogenic emergency situations* (Moscow: KolosS) 520 p
[5] Shannon K 1963 *Mathematical theory of communication* Works on information and cybernetics (Moscow: Mir) pp 233-332
[6] Zolotova E V 2016 *Cadastre fundamentals: Territorial information systems: Textbook for universities* (Moscow: Fund "Mir", Acad. Project) 416 p
[7] Brillinger D 2016 *Time series. Data processing and theory* (Moscow) 103 p
[8] Dem'yanov V V, Saveliev Ye A; Harutyunyan R V (ed) 2010 *Geostatistics: theory and practice* (Moscow: Nauka) 327 p
[9] Haken G 1985 *Synergetics. Hierarchies of instabilities in self-organizing systems and devices* (Moscow: Mir) 420 p
[10] Solovyov I A 2008 Prediction of ecological disasters based on the analysis of statistical information Coll. Ecological problems of regional environmental monitoring (Moscow: Georesurs) pp 36-41
[11] Baksansky O E 2019 *Modeling in science: Building physical models* (Moscow: Lenand) 160 p