IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 4, Ser. II (Jul. –Aug. 2022), PP 63-67
www.iosrjournals.org

Uniform Representation of Data for the Eco Monitoring Information System

David Kobrava, Manana Khachidze
(Computer Sciences Department, Iv.Javakhishvili Tbilisi State University, Georgia)

Abstract:
Background*: Environmental monitoring remains one of the important problems of the modern world, in the management of which information systems play an important role. Raw data received from various sensor devices need to be processed with appropriate methods and techniques. The variety of sensor devices used for monitoring generates a variety of (non-uniform in format) data streams. This, in turn, complicates the analysis process.*
Materials and Methods*: The method of analytical heuristics is used to achieve a uniform presentation of data. This method allows all data types to be represented in the same format. Many data analysis techniques can be used to process homogeneous data. including a description of the state of the environment in the form of concepts.*
Results*: As a result of the above we have obtained a method (a language) which enables us to: Bring various parameters to a single (uniform) system of measurement; Describe classes of water pollution as concepts, which allows determination of the quality of surface water and assessment of the condition of a water body.*
Key Word*: Ecomonitoring; Data Uniform; Concept; Data Analyses*

Date of Submission: 15-08-2022 Date of Acceptance: 31-08-2022

I. Introduction

Proceeding from the general definition of ecomonitoring as an information system that covers observation, assessment and forecasting of changes in the condition of the environment, it is natural to conclude that the main task as assigned to the above system is accumulation and systematization of data about the state of the environment [1]. This data is to include data concerning both geographical and geological aspects of the state of the ecosystem, which involves a wide range of investigative methods and techniques.Data sources are mainly represented by various sensor devices. Accordingly, the volume of data increases, in result we are getting in big data[2]. Understandably, ecologic information is quite varied and requires elaboration of a single pattern of data presenting that will facilitate storage and subsequent analysis of these data with a view to assessing and forecasting the condition in which the environment finds and is likely to find itself in the future. To that end, we have set ourselves the task of working out a single methodology for presenting and processing the data on the basis of ecologic standardization and the method of analytical heuristics.Data presented in a unified form will enhance the use of modern data analysis tools, especially in the case of big data.

II. Parameters and Ways of Their Representation

To be more precise, let us opt for monitoring the condition of surface water, where the chief indicator is the quality of water and where the parameters of its assessment fall into the following groups [3]:

1. general physical and chemical indicators;
2. organoleptic indicators, or parameters that immediately affect the senses of a human being (smell, touch, eyesight);
3. the main inorganic substances;
4. the main organic substances;
5. bacteriologic and parasitological substances;
6. disinfectants and products of disinfection;
7. radiological indicators.

In some countries the main indicator of the degree of pollution of the water under testing is the maximum permissible concentration (MPC) of a pollutant in a unit volume of water, of mass, etc. that does not impair the health of a living organism in any practical way [3]. In case of concentration of the pollutant$C_i \leq MPC$, the norm is fixed from parameter*i*, i.e. according to this indicator, the condition of water is satisfactory. With $C_i > MPC$the degree of pollution of the water under testing is fixed from parameter*i, Cι*is the concentration of substance *i*, $i = 1 \dots \dots n$ – is the number of parameters under consideration. Depending upon the extent to which the concentration of the pollutant exceeds the MPC, the degree of pollution of the water under test containing that pollutant is classified as moderate, heavy and extremely severe.

To bring all the parameters to a single system of measurement in relative units, it is hereby proposed to express their amounts in terms of the corresponding MPCs or analogous values equivalent to their MPCs:

$$N_i = \frac{C_i}{MPC_i},$$

Where:

C_i is the concentration of substance *i* in the water,

$N_i$is the value of parameter *i* expressed in relative MPC units, $i = 1. \dots \dots n.$

To determine the quality of the water under testing according to any of the above parameters it is sufficient to check the condition that is similar for all the parameters and see whether $N_i \leq 1$. If it is so, then the normalcy is fixed according to *i* parameter, i.e. according to this parameter, the quality of the water under testing is satisfactory, but if $N_i > 1$, the degree of pollution is fixed according to *i* parameter, whereN_i – is the value of parameter $_i$ expressed in relative units, $i = 1 \dots \dots n$ (Fig.1).

However, there are some parameters where the notion of maximum permissible concentration is not applicable. These are mainly the parameters from the group of general physico-chemical and organoleptic indicators of water quality. In these parameters the MPC is replaced with the so-called norms of quality that fix the upper or the lower limit or interval of permissible values, for instance, mineralization of water, the oxygen that is dissolved in water and the acidity indicator or the pH whose norms of quality are shown in Table 1.

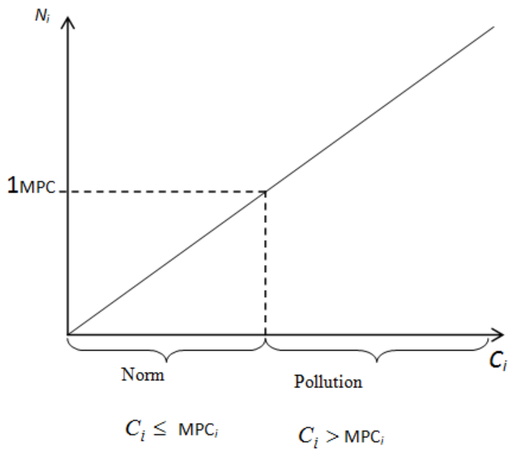

Figure.1. Dependence of parameter *i* expressed in relative units and the concentration of the *i* substance

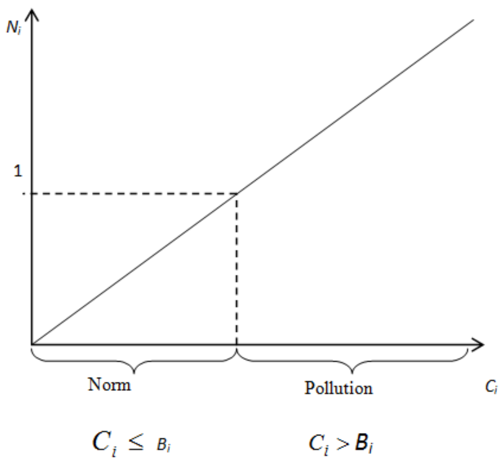

Figure 2. Dependence between values N_i and C_i in the case where the norm of quality fixes the upper limit of permissible values of parameter *i*.

In the case where the norm of quality fixes the upper limit $B_i$of permissible values, the corresponding parameter will be transferred into the relative units, like in the case of MPC:

$$N_i = \frac{C_i}{B_i},$$

where:

C_i is the value of the i parameter,

N_i is the value of the i parameter expressed in relative units;

$B_i$is the upper limit of permissible values of parameter *i*, $i = 1. \dots \dots n$.

And in this case the condition for determining the quality of water remains the same, if $N_i \leq 1$, the norm is fixed according to *i* parameter;if $N_i > 1$, the degree of pollution is fixed according to *i* parameter,where $N_i$is the value of the *i* parameter expressed in relative units, $i = 1. \dots \dots n$ (Fig.2)

TABLE NO 1: Parameters for which there are norms of quality that fix the upper or the lower

Indicator	Norm
Mineralization in mg/l	≤ 1000
Dissolved oxygen mg O2/L	> 6
Acidity indicator	6.5 – 8.5

In case where the norm of quality fixes the lower limit A_i of the permissible values, the relative values of the corresponding parameter are calculated as under:

$$N_i = \frac{A_i}{C_i}$$

where:

Ci is the value of the *i* parameter,

Ni is the value of the *i* parameter expressed in relative units,

Ai is the lower limit of permissible values of the*i* parameter, $i = 1. \ldots\ldots n.$

III. Ease of Use Description of the Method

The formula for expressing the values in relative units was created on the basis of the following considerations: like in the above cases, the relative values among the multitude of permissible values of the parameter are within the limit from 0 to 1; they reach 1 on the limit of $A_i$and assume values that exceed 1 and are beyond the interval of permissible values (Fig.3).

But when the norm of quality sets the interval of permissible values $[A_i;B_i]$, i.e. when with $A_i \le C_i \le B_i$, the norm is fixed according to parameter *i*;

when $C_i < A_i$ or $C_i > A_i$ the degree of pollution is fixed according to parameter *i*,

where C_i is the concentration of the *i* substance, then we use a combination of the two previous approaches, namely: in the interval (A_i, *Bi*) *Ni* does not exceed 1, i.e. we are not concerned with its concrete value, we express it as a length of a straight line that is parallel to (Ai, Bi) with ordinate "1". To the left of $A_i$there is a hyperbolic curve (like in case "2"), while to the right there is a straight line that passes across point *Bi*and the origin of coordinates (like in case "1"). Thus, the relative values of the corresponding parameter are calculated as under:

$$\begin{aligned} &if\ C_i < A_i, then\ N_i = \tfrac{A_i}{C_i};\\ &if\ A_i \le C_i \le B_i, then\ N_i = 1;\\ &if\ C_i > B_i, then\ N_i = \frac{C_i}{B_i}. \end{aligned}$$

The condition for determining the quality of water from the relative value of the parameter of the interval remains the same (Fig.4).

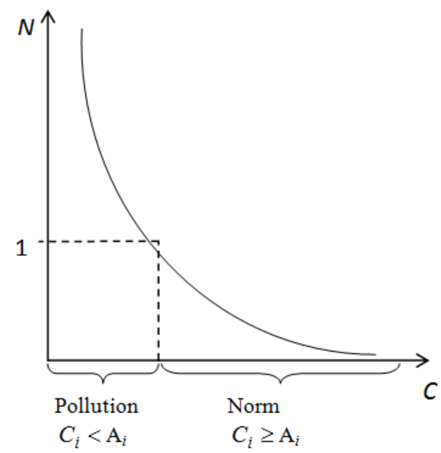

Figure 3. Dependence between the values of $N_i$and $C_i$in case when the norm of quality fixes the lower limit of permissible values of parameter *i*.

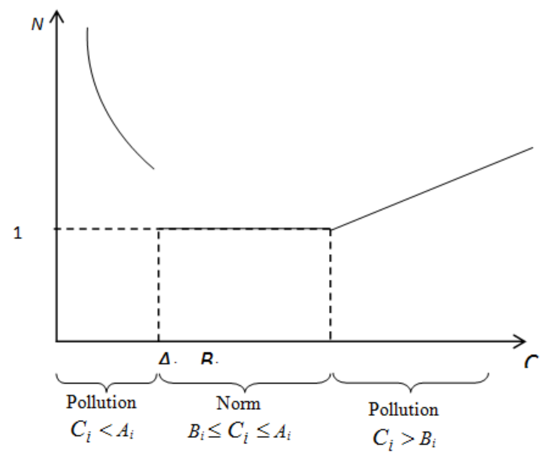

Figure 4. Dependence between the values of N_i *and* $C_i$in case when the norm of quality fixes the interval of permissible values of parameter *i*.

After expressing all the parameters in relative units, the condition of the water body is described within the system of parameters of a similar type, and pollution can be marked as soon as the relative value of this indicator has exceeded 1. As mentioned earlier, different degrees ofexcess of the norm mark different levels of pollution, namely [3]:

$N_i \le 1$ – permissible concentration

$1 < N_i \le 4$ – moderate pollution

$4 < N_i \le 10$ – heavy pollution

$10 < N_i \leq 100$ – extremely severe pollution.

Finally, the condition of a water body can be described within the system of four-digit linguistic markers.

The next stage that follows description of the condition of water bodies is analysis of all this information with a view to assessing the quality of the water they contain. The best technique to be used here is the conceptual technique of forming a notion of analytical heuristics, which allows appropriately assessed "trajectories" or descriptions in terms of "sign x value" obtained at the point of entry to be run through a system of filters and at the exit yield the concept of a corresponding notion in terms of minimized disjunctive normal form (DNF) of the condition vectors composed of a multitude of values of the signs [4,5].

And indeed, if we consider as "trajectories" the condition of water bodies the quality of whose water will be described by a cluster of parameters expressed in relative units, and the multitude of these bodies will be broken down into classes depending on the degree of water pollution, then we shall obtain concepts of production rules that will allow us to place this or that water body in this or that class, which means that we will thereby assess its condition.

Another merit of the above conceptual method is its universal nature from the point of view of its applicability to all types of data: binary and multivalued ones, discrete and continued ones, numerical and linguistic ones. The catholicity or universality of this method stems from the vectors of condition that describe the state an object is in and serve as a basis for the concept of assessment of this state. The formation of these vectors for each type of signs and each level of the concept is considered in detail in [6].

It ought to be mentioned that the condition of an object can be described not only by vectors of condition, for each level of precision determines its own record of the object's condition. Transition to a still higher level is to be resorted to when the selected concept fails to unequivocally classify objects of educational sampling. Generally speaking, an object may contain vectors of condition at various levels [7].

In our case, when the condition of the water body is represented through four-digit linguistic signs, the condition vectors will look as under:

$\psi_i^{(I)} = (1,2,\bar{3},\bar{4})^T$, $\overline{\psi_i^{(I)}} = (\bar{1},\bar{2},3,4)^T$ - vectors of condition at the *I* level

$\psi_i^{(II)} = (1,\bar{2},3,\bar{4})^T$, $\overline{\psi_i^{(II)}} = (\bar{1},2,\bar{3},4)^T$ - vectors of condition at the II level,

where numbers 1, 2, 3, 4 re-indicate the linguistic meaning of the concepts of "permissible concentration", "moderate pollution", "heavy pollution" and "extremely severe pollution" respectively. If we move over to the original numerical values, the vectors of condition will correspond to the following intervals:

$\psi_i^{(I)} - [0;4]$ $\quad \overline{\psi_i^{(I)}} - [4;100]$,

$\psi_i^{(II)} - [0;1] \cup [4;10]$, $\quad \overline{\psi_i^{(II)}} - [1;4] \cup [10;100]$,

which means that each linguistic meaning can be obtained as a conjunction of vectors of condition at the I and the II levels:

$N_i \leq 1$is permissible concentration - $\psi_i^{(I)} \& \psi_i^{(II)}$;

$1 < N_i \leq 4$is moderate pollution - $\psi_i^{(I)} \& \overline{\psi_i^{(II)}}$;

$4 < N_i \leq 10$is a heavy pollution - $\overline{\psi_i^{(I)}} \& \psi_i^{(II)}$;

$10 < N_i \leq 100$is an extremely severe pollution - $\overline{\psi_i^{(I)}} \& \overline{\psi_i^{(II)}}$.

Thus, each value of the *i* sign can be unequivocally represented both in the "trajectories" and in the concept as a conjunction of vectors of the condition at the I and the II levels, and any further adjustment of the concept will be unnecessary.

IV. Conclusion

As a result of the above we have obtained a method (a language) which enables us to:

1. Bring various parameters to a single (uniform) system of measurement;
2. Describe classes of water pollution as concepts, which allows determination of the quality of surface water and assessment of the condition of a water body.

The data presented in a uniform are raw data collected from various sensors used in environmental monitoring. Given different observation time intervals, these data increase quantitatively to such an extent that it is most convenient to use data analysis tools and methods adapted to big data processing.

The description of various levels of water pollution presented in the form of concepts can be used in machine learning as part of the main functionality of the environmental monitoring information system.

References

[1]. El-Gayar, Omar and Fritz, Brian D. (2006) "Environmental Management Information Systems (EMIS) for Sustainable Development: A Conceptual Overview," Communications of the Association for Information Systems: Vol. 17 , Article 34. DOI: 10.17705/1CAIS.01734 Available at: https://aisel.aisnet.org/cais/vol17/iss1/34

[2]. Lucivero, F. Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives. Sci Eng Ethics 26, 1009–1030 (2020). https://doi.org/10.1007/s11948-019-00171-7

[3]. Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda, Geneva, 2022. https://www.who.int/publications/i/item/9789240045064

[4]. Chavchanidze V. V. Analiticheskiye evristiki iskussnvennogo intellekta pri formirovanii ponyatiy, opoznanii obrazov i klassifikatsii obyektov. VINITI. № 2080-70 Dep. (Analytical Heuristics of the Artificial Intellect in the Formation of Notions, Identification of Images and Classification of Objects, VINITI, #2080-70 Dep.

[5]. Chavchanidze V. V. Analiticheskoye Resheniye Zadachi Formirovaniya Ponyatyi i Raspoznavaniya Obrazov (Analytical Resolution of the Problem of Identification of Notions and Recognition of Images). Proceedings of the Academy of Sciences of the Georgian SSR, vol.63, No.1, 1971.

[6]. M. Khachidze, M. Mikeladze. Data Unification Algorithm for Representing Incomplete and Indefinite Information in the Medical Expert System. Proceedings of the EUROPEAN COMPUTING CONFERENCE (ECC'09), Tbilisi, 2009, pp.348-352.

[7]. Chavchanidze V. V., Korneeva A.V. Analiticheskiy Filtratsionnyi Metod Formirovanoya Poynatiy (Analytical Filtration Method of Formation of Notions, Proceedings of the Academy of Sciences of the Georgian SSR, vol.65, No.3, 1972).

[8]. Carsten Dormann. Environmental Data Analysis. 2020.ISBN : 978-3-030-55019-6

[9]. Hino, M., Benami, E. & Brooks, N. Machine learning for environmental monitoring. Nat Sustain 1, 583–588 (2018). https://doi.org/10.1038/s41893-018-0142-9

David Kobrava, et. al. "Uniform Representation of Data for the Eco Monitoring Information System." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 24(4), 2022, pp. 63-67.