Prezentat la 1st INTERNATIONAL U.O.C. – B.EN.A. conference „The sustainability of pharmaceutical, medical and ecological education and research – SPHAMEER - 2013”, under section **Environmental Pollution (air - water - soil)**

POLLUTION PREDICTION METHODOLOGY FOR SITE IMPACT/RISK ASSESSMENT USING ROUGH SET THORY

Madalina Arama[1*],

[1]INCD-ECOIND Bucharest
Drumul Podu Dambovitei 71-73, Sect. 6, Cod postal 060652;
tel: 04.021.410.67.16 / 410.03.77 / 410.24.01
fax: 04.021.410.05.75 / 412.00.42;
email: ecoind@incdecoind.ro; http: //www.incdecoind.ro
**madalinaarama@yahoo.com*

The paper presents the methodology for environmental risk assessment using RST that analysis the probability and the gravity of consequences of exposure of natural environmental components (air, surface water, soil, sub-soil, ground water) when the hazard of significant pollution act on them considered as targets. The basic elements of the RST are presented in relation with their potential use in environmental impact/risk assessment for the continual improvement of the environmental performance to any organization.

The new proposed methodology for site environmental impact/risk assessment brings a new RST based decision table with a set of certain and transparent decision rules in order to conform the current environmental legislation and to allow classification of possible environmental pollution states in the analyzed site named observations/events in nine risk classes scored increasingly from 1 to 9. Each risk class is characterized in order to be user friendly in practical cases where this decision table can serve as reference.

Key words: RST, environmental impact/risk assessment methodology

1. Introduction

The new methodology brings a decision frame with transparent decision rules based on Rough Set Theory and current environmental concepts in the field. The environmental adverse impact analysis and assessment and its associated risk are essential endeavors that any organization are pursuing in relation to the decision regarding the adequate management of its environmental emissions. The proposed methodology will improve the current existence environmental impact/risk methodology based on scientific argumentation. The large accepted definition of risk assessment is the process of estimation of possibilities/probabilities that a certain particular event occurs in certain circumstances. In the context of a sustainable development measuring and managing the present or future pollution is a must for any organisation that has significant environmental aspects. Having an appropriate model in the environmental field for predicting the future adverse effects and the associated risk become a necessity. „Models' predictions cannot be better than the quality of their input data. The data reconciliation includes the detection of non-conforming data, their isolation, the identification of causes and the preparation of data set according to the modeled objective. The data mining techniques can help to improve the model predictions by considering the each mentioned problems"[1] .The new methodology fulfills the necessity to supplement the existing environmental support decision methodologies in uncertainty conditions based on probability theory with new approaches. They add to the pur probabilistic methods that consider the only measurements uncertainty the ones that take into account also the epistemic uncertainty. Among those last methods is the initial Rough Set Theory and its probabilistic new development frame aproaches [2,3,4]. „Combined with complementary concepts like fuzyy sets, the statistics and logic data analysis, rough sets have been exploited in hybride approaches to improve the data analysis tools performance"[5,6] „Among the data mining thechniques, the Rough Set Theory, proposed by Pawlak is a mathmatical instrument and an appropriate approach to face the imprecision, lack of completness and uncertainty in the anallysed data"[1,7,8].

2. General Background of RST Methodology

Because pollution phenomena is pervaded both by aleatory and epistemic uncertainty, the RST can be the appropriate tool for handling them. It can describe the observations/events characterized by incomplete information using vague concepts expressed in natural language like "big", "small", "not to big", "not to samll". The majority definitions recomended by environmental laws that try to classify pollution are using this kind of language. To classify an environmental state under a pollution class defined this way, the use of RST seems to be especially desirable because RST is able to handle such vague concept. RST uses to express vague concept a set named rough set. In a rough set there are objects that can certainly be classified as part of a concept/set and others that can possibly be classified as being part of that concept/set. To a such rough set, RST associates two crisp/well defined/classical sets a lower approximation set having events that are certaintly part of the concept/set and an upper aprpoximation set having events that possible are part of the concept/set. A rough set has a non-empty boundry region where the events cannot be uniquely classified neither as being part of the concept nor as being part of its complement [9].

3. RST site assessment methodology

In the present methodology the risk is a function of hazard and favourable conditions. The main favourable condition for a negative consequence to occur at certain targets is their exposure. A target is exposed when the hazard reaches it. The hazardous situation from the polluted sites, actioning on the environmental components (air, water, soil, sub-soil etc) is considered when one or more pollutants concentrations exceed maximum allowable concentration abraviated as MAC ($C_{poluant}$ >MAC). When those pollutants' concentrations greater than MAC reach the targets, the premises for the occurence and development of different adverse effects of smaller or greater serverity in those targets, in a certain specified period of time, are fulfilled. The severity of the produced consequences/damages increases with the hazardous pollutant(s) properties, with the exposure time (the time of contact $C_{poluantului}$> CMA on the target) and with the conditions that favor the pollutants to reach the target and to accumulate upon it. The present environmental risk assesment methodology based on RST proposes a decision table for the calssification of events representing possible environmental pollution perceived/measured states in nine risk classes ranked with grades from 1 to 9. The pollution states are denoted $X_1 \div X_{13}$ and they are characterized by a set of three conditions attributes denoted H,T,F of generic character and one decision attribute denoted R reprezenting the risk decision classes.

a. Attribute condition H (abraviated H from hazard. It represents the degree of hazard coming from one or more pollutants and takes values from the set {0,1,2,3} as follows: "0" when the pollutant is not hazardous and "1","2", or "3" when the pollutant has hazardous properties;

b. Attribute condition T(abraviated T from target). It represents the fact that the target is reached or not by a pollutant having a concentration greater than MAC in all relevant proposed casses of investigation. It can take any values from the set {0,1}as follows: "0" when pollutant does not reach the target and value "1" when pollutant does reach the target;

c. Attribute condition F (abreviated F from favorable): The attribute F represents relevant favourable conditions that favour pollutant(s) to reach the target and can take values from the set {0,1} as follows: value " 0" when does not exist favourable relevant conditions and value " 1 " when there are favorable relevant conditions;

4. Attribute decision R (abraviated R from risk): It represents the risk class that can take values from the set {1, 2, 3, 4, 5, 6, 7, 8, 9} as it follows in the Decision Table. In the acceptance of the current method those values express the monotonic increase of the risk from 1 to 9 by increasing the premises, i.e. the possibilities in classes R=1÷3 or of probabilities in classes R= 4÷9 that different negative consequences of monotonic severity increasing upon considered target. The decision rules are certain from the RST point of view and set according to the legal environmental international requirements. For example, if the environmental pollution observations/states are of type X_1 (Condition attribute H=0, T=0, F=0) then the decision is R1 etc. The table 1 corresponds to the pollution sates when on the site one or more pollutants having $C_{polluant}$ > MAC, exist in the respective environmental compartment (air, surface water, soil, sub-soil, ground water).

Table 1 – Decision Table for Site Impact/Risk Assessment

Obs/ Events	H	T	F	Risk Class-R	Risk Class-R Description	Comments of risk classes use (increasing risk from up to down)
0	**1**	**2**	**3**	**5**	**6**	**7**
X1	0	0	0	**1**	**very, very low**	-Pollutant is not hazardous (so H=0). -Pollution is significant (> MAC) with this type of non-hazardous pollutant (with H=0), it does not reach the considered target in the medium (so T=0), however, the pollution possibility to reach the target in the future cannot be excluded. -There are not in present favorable conditions so that the hazard of significant pollution to reach and act upon the target (so F=0); **-Risk Class R=1; Risk is "very, very low "; it is a potential risk** because the pollutant is not hazardous (H = 0); it exceeds the MAC in the medium where it is measured but it does not reach the target so, the probability that the hazard *($C_{pollutant(s)}$ >MAC)* ***to act upon the target is 0.*** **-The Risk Class R=1** ***shows that in this type of considered observations/events the probability of occurrence of negative consequences at the target is 0 because the target is not reached (T=0).***
The tables goes in the same way from class risk R=1 (H=0, T=0,F=0) through class risk R=5 (H=1,T=1,F=1) reaching class risk R=9 (H=3, T=1,F=1) as pollutants/group of polutants has increasingly hazardous properties and reaches the target in conditions favorable for pollution concentration mganification at the target. 1. Class risk R=1 (very, very low), 2. Class risk R=2 (very low), 3. Class risk R=3 (low), 4. Class risk R=4 (low towards medium) 5. Class risk R=5 (medium) 6. Class risk R=6 (medium towards high), 7. Class risk R=7 (high), 8. Class risk R= 8 (very high), 9. Class risk R=9 (very,very high)						
X13	3	1	1	9	**very,very high**	-Pollutant is very hazardous (so H=3). -Pollution is significant (>MAC) with this type of high hazardous pollutant (with H=3), it does reach the target (so T=1). -There are in present favorable conditions that pollutant concentration to magnify at the target F=1. **-The Risk Class R=9; Risk is "very, very high" ; it is a real risk** because it is a highly hazardous pollutant (with H =3), it exceeds the MAC in the medium where it is measured and does reach the target, so , the probability that hazard ($C_{pollutant}$ > MAC) magnifies at the target is 1; **-The Risk Class R=9 shows that this type of considered observations/events,** ***the probability of occurrence of negative consequences because the pollutant/group of pollutant despite the fact that has the same degree of hazard (H=3), the target is reached (T=1) like in risk class R=8 however, in addition there are also favorable conditions for the magnification of concentration at the target because F=1.***

Conclusions

The advantages of the new proposed methodology are: 1) it offers a rapid and easy way to implement computation support for data selection, reduction and data analysis; 2) it takes into account the complex multicriteria nature of managing pollution emissions; 3) It can manage simultaneously the aleatory and epistemic uncertainty being tolerant to the vague concepts as is the case of the majority of environmental/impact risk assessment information; 4) it uses hybrid quantitative-qualitative input data that are processed according to the rough set theory capabilities 5) it ofers a transparent frame to support the decision in uncertainty conditions.

References

1. L. FENG: *Water Treatment Plant Decision Making Using Rough Multiple Classifier System.* Procedia Environmental Sciences, 11, 1419 -1423 (2011).

2. H. ZHANG, J. ZHOU, D. MIAO, C. GAO: *Bayesian rough set model: A further investigatio.*, International Journal of Approximate Reasoning, 53, 541-557(2012).

3. Y. YAO: *Probabilistic rough set approaximations.* International Journal of Approximate Reasoning, 49, 255-271 (2008).

4. D. SLEZAK, W. ZIARKO: *The investigation of the Bayesian rough set model*. International Journal of Approximate Reasoning , 40, 81-91 (2005).

5. G. LIU, W. ZHU: *The algebraic structures of generalized rough set theory*. Information Science, 178, 4105-4113 (2008).

6. Z. PAWLAK, A. SKOWRON: *Rough sets and Boolean reasoning*. Information Sciences 177, 42-73(2007).

7.X. ZHANG, Z. MO, F. XIONG, W. CHENG: *Comparative study of variable precision set model and graded rough set model*.International Journal of Approximate Reasoning 53, 104-116 (2012).

8.M.REBOLLEDO:*Rough intervals-enhancing intervals for qualitative modeling of technical systems.* Artificial Intelligence, 170, 667-685(2006).

9. Z. PAWLAK, A. SKOWRON:*Rudiments of rough sets*. Information Sciences 177, 3-27 (2007).