A Knowledge-Based Environmental Protection System

Mihaela Oprea

University of Ploiesti, Department of Informatics, Ploiesti, 2000, Romania
Email: mihaela@upg-ploiesti.ro

Abstract. Making decisions in the field of environmental protection management is highly complex and a solution for reducing this complexity is given by the knowledge-based approach. Such a solution could structure the domain expert knowledge into a knowledge base and later will use this knowledge during the reasoning process. The paper describes a knowledge-based environmental protection system designed for air pollution control in urban regions.

1 Introduction

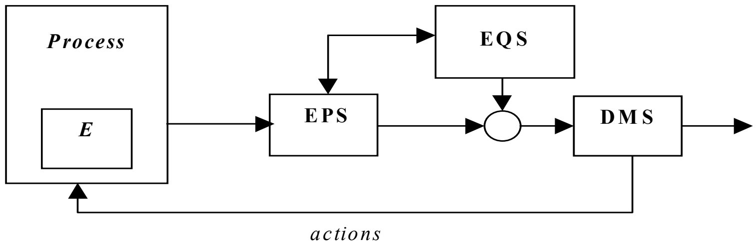
The high complexity of making decisions in the field of environmental protection management derives from a lot of factors, environmental, social, economical and so on, that should be analyzed and which have in almost all cases contradictory interests [1]. The use of a knowledge-based approach is a solution for reducing this complexity, the environmental protection system being modeled as a knowledge-based system (KBS) [8, 9]. In this paper we focus on the atmospheric environmental protection in urban regions. Air pollution comes from many different sources: stationary sources such as factories, power plants, and smelters and small sources such as dry cleaners and degreasing operations; mobile sources such as cars, buses, planes, trains, trucks; and naturally occurring sources such as windblown dust, all contribute to air pollution. Air quality can be affected in many ways by the pollution emitted from these sources and a wide variety of pollutants could be emitted by these sources. The main role of an atmospheric protection system is the reduction of the pollutants emissions in order to reduce public health risks and to protect sensitive ecosystems. We have developed a knowledge-based system whose aim is to provide expert assistance to atmospheric environmental management in the area of Ploiesti, an industrial town where the predominant industry is the petrochemical one.

The paper is structured as follows. Section 2 describes the environmental protection system (EPS), focusing on the atmospheric environmental subsystem. Section 3 presents the architecture of the designed knowledge-based system, SBC_ProtMed. A brief discussion about the experimental results obtained so far is made in section 4. Finally, in section 5 we conclude the paper.

2 The Environmental Protection System

In order to solve the environmental problem the experts need to use not only a wide range of environmental data, but also specific knowledge about the environmental process (considered as a global process) in order to assist the decision making system during monitoring, diagnosis, forecasting and improvement of the environmental quality [2]. The purpose of an environmental protection system is to predict the process behavior and to check the admissible limits of the observed parameters (e.g. concentrations of air pollutants) according to the existing environmental quality standards. Figure 1 presents the block diagram of the environmental protection control system. The feedback from the decision making system will generate different actions applied on the environmental process, trying to maintain the analyzed parameters between predefined limits. In our case, of an atmospheric protection system, the main actions are emission reductions that would result in major reductions in the concentrations of atmospherically transported pollutants. Another important aspect regarding the designed EPS is that a knowledge-based model would allow including the real time capability.

The quantitative information derived and integrated from measurements and observations will tell the decision maker how much, how many and how fast changes are occurring and whether the existing standards are fulfilled [4]. In the case of an atmospheric EPS, the quantitative information is given by the concentrations of air pollutant substances measured in different sites of the environmental controlled area. For this purpose, the EPS has an atmospheric monitoring subsystem whose role is to be aware of all levels of pollutants in these sites. In urban regions, due to the industrial activities and to the exhaust gas from different types of vehicles (cars, buses, trains, etc), there are emitted in the air a series of pollutants, whose concentrations should be monitored and kept in the suboptimality intervals. The quantity and the composition of the emitted substances depends on many factors, in particular on the the industrial branch, on the characteristics and quantities of the used base production materials, on the type and quantity of the fuel used, on the applied technology and on the effective measures of atmospheric environmental protection. Also, of great importance are the meteorological factors that could increase or decrease the pollution level of the emitted pollutants. In the case of the Ploiesti area, the main air pollutants substances are emitted due to the petrochemical industry, the town being surrounded by a series of oil refineries and other chemical plants. The existing air monitoring subsystem measures and analyses mainly the following substances: NH_3, NO_2 (NO_x), SO_2, fenol, H2S, SO_2 (2-), CO, suspended particulates emissions (respirable or fine) etc. The concentrations of these pollutants are measured in various locations over the area of Ploiesti. As some of these pollutants are contributions from motor vehicles, some monitored locations focus on the roadside, street-level concentrations. The environmental network topology allows to collect the quantitative information from all workstations, either on line, either in situ and to coordinate the decisions regarding preventive measures in critical situations through a supervisor system. The designed knowledge-based atmospheric EPS is located on the supervisor system and so far, could deliver warnings for the sites with potential atmospheric pollution risk on the basis of the current and past measurements of the pollutant substances concentrations and of the meteorological data, providing possible solutions for avoiding such situations.



E – Environment
EPS – Environmental Protection System
EQS – Environmental Quality Standards
DMS – Decision Making System

Fig. 1. The block diagram of the control of an EPS

A common characteristic of the environment-related tasks is that they rely heavily on climate data. Therefore, meteorological data are of great importance as the temperature of the air, the speed of the wind, the degree of solar radiation, the rainfalls, etc, will influence the degree of the atmospheric pollution. In order to solve environmental forecasting and planning problems some meteorological predictions are usually needed. As a consequence, there should be a data exchange between the environmental network and the meteorological network. Figure 2 presents the block diagram of the atmospheric EPS (AEPS). Both systems, environmental and meteorological are represented in the figure. The subsystems of the AEPS are: the atmospheric environmental forecasting system (AEFS), the atmospheric environmental modeling and representing system (AEMS), and the knowledge-based system (KBS). AEFS is in our implementation a feed forward artificial neural network that makes predictions regarding the evolution of the pollutants concentrations based on past measurements [7]. The AEMS includes diffusion models (e.g. the Gaussian model) for air pollution substances, different models for representing environmental data (e.g. time series, images, graphics, tables), simulation and control models etc. All the environmental data are measured and observed from the environment and are collected in the databases DB_E of the AEPS. The meteorological system is composed by the following subsystems: the meteorological forecasting system (MFS), the meteorological modeling and representing system (MMS). All the meteorological data are measured and observed from the environment and are collected in the databases DB_M of the meteorological system.

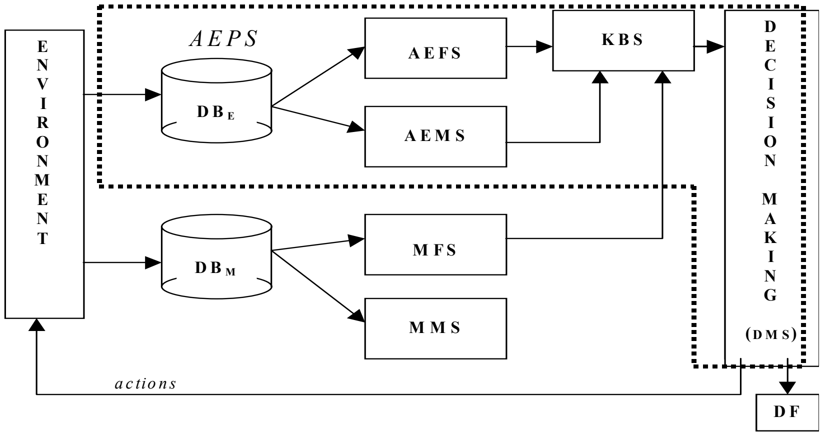


Fig. 2. The block diagram of AEPS

The decision making system (DMS) has to solve environmental protection tasks such as diagnosis, forecasting, planning and designing [6]. The system receives information from the knowledge-based system and after analyzing them will take some decisions that could be applied to the atmospheric environmental protection for the current urban region. The decisions will be communicated also to other decision factors (DF).

3 The Architecture of *SBC_ProtMed*

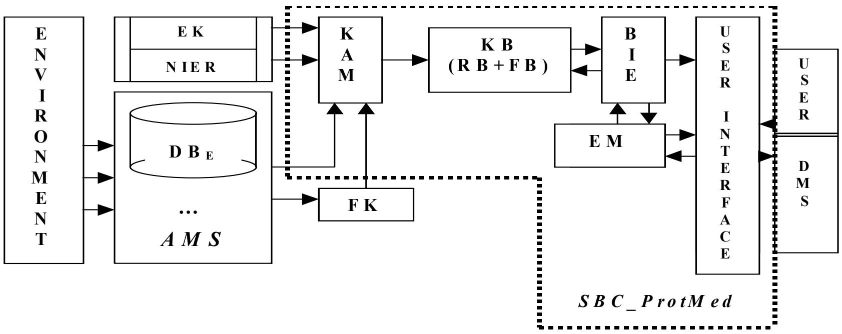
Figure 3 shows the architecture of the knowledge-based system *SBC_ProtMed*. The main role of this KBS is to provide ***AEPS control knowledge*** to the decision making system (DMS). The *knowledge acquisition module* (KAM) takes the knowledge from different sources, and then filter and structure them in a correctly and efficiently manner in order to introduce them in the knowledge base. The principal knowledge sources are expert knowledge, data from the atmospheric monitoring system, national and international environmental regulations (e.g. laws, conventions, agreements, national and international air quality standards), forecasting knowledge (meteorological and air pollution predictions).

The knowledge is represented under the form of production rules and the general form of a rule is the following:

RULE <label>
IF <premise> **THEN** <conclusion> **CNF** <numerical_value>;

where, <premise> is <condition_1> **AND**/**OR** <condition_2> **AND**/**OR** … **AND**/**OR** <condition_n>, <conclusion> is either an action or a new knowledge (deduced by inference), **CNF** is the confidence factor that gives the degree of confidence in the truth of the rule; CNF has a numerical value in the interval [0, 100].

Some of the rules from the rules base (RB) were generated by using inductive learning algorithms [10]. The knowledge acquisition process is a long and very difficult step in the system implementation due to the lack of a well structured expert knowledge in the field of environmental protection, as well as to the incomplete data and uncertainty that is involved by environmental processes, and also to the high complexity of the whole environmental system. The quality of knowledge sources is of great importance for the utility of the KBS to an effective environmental protection. The inference engine uses a backward chaining of the rules and includes a module for the management of uncertainty. As in the atmospheric environmental protection we deal with incomplete knowledge or even missing data, we have introduced uncertainty in the knowledge base by using confidence factors [5]. Each rule has associated a confidence factor (called CNF) that is given by the experts in atmospheric environmental protection.



AMS – Atmospheric Monitoring System
EK – Expert Knowledge
NIER – National & International Environmental Regulations
KAM – Knowledge Acquisition Module
KB – Knowledge Base (RB – Rules Base, FB – Facts Base)
BIE – Backward Inference Engine
EM – Explanatory Module
FK – Forecasting Knowledge
DMS – Decision Making System

Fig. 3. The architecture of *SBC_ProtMed*

The inference engine (module BIE in figure 3) will manage the knowledge uncertainty by using the following rules:

1. CNF(conclusion) = CNF(premise)×CNF(rule)
2. $CNF_{R1,\,R2}{}^{p} = CNF_{R1}{}^{p} + CNF_{R2}{}^{p} - CNF_{R1}{}^{p} \times CNF_{R2}{}^{p}$
3. CNF(premise) = min (CNF(condition$_i$)), where i=1, …, n

The first rule gives the confidence factor in the conclusion of a rule when the confidence factor in the premise of the rule and the confidence factor in the rule are known. Rule 2 determine the confidence factor in a conclusion p that is encountered in two different rules R_1 and R_2. Rule 3 is applied in the case of rules that have a premise composed by several conditions connected by the AND logic connector. In the case of a premise composed by several conditions connected by the OR logic connector rule 2 is applied.

The facts base contains the permanent facts and the facts of the current context, i.e. the initial facts and the deduced facts. The form of a fact is the following:

<variable> <operator> <value>

where the value of the variable could be symbolic or numerical, and an operator could be a relational operator from the set {<, <=, >, >=, =}.

The facts are taken mainly from the databases of the atmospheric monitoring system. We have used three types of facts: numeric, boolean and fuzzy. The fuzzy interval (similar with that used in [3]) includes a set of nine ordered values: impossible, almost impossible, slightly possible, moderately possible, possible, quite possible, very possible, almost sure, and sure. For the first stage of our experiments we have associated certainty factors to these fuzzy values [11]. The certainty values are linguistic terms defined by the environmental protection expert. The internal representation of each term is a fuzzy number in the interval [0, 1] and allows dealing with uncertain facts and with rules whose uncertainty concerns the strength of the implication. For computational reasons this representation is parameterized.

The knowledge-based system allows the expert to define the term set of linguistic certainty values which constitutes the verbal scale that he/she and the users will use to express their degree of confidence in the rules and facts respectively. Each linguistic value is represented internally by a fuzzy interval (a fuzzy number) i.e., the membership function of a fuzzy set on the real line (i.e. on the truth space represented by the interval [0, 1]). The membership functions can be interpreted as the meanings of the terms in the term set.

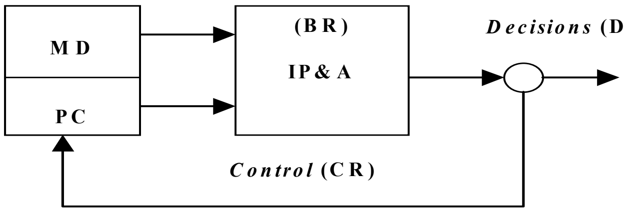
TERM_SET = {*impossible, almost impossible, slightly possible, moderately possible, possible, quite possible, very possible, almost sure, sure*}

The parametric representation of each term:
impossible = (0, 0, 0, 0)
almost impossible = (0 , 0, 0.04, 0.07)
slightly possible = (0.04, 0.06, 0.15, 0.17)
moderately possible = (0.10, 0.16, 0.35, 0.45)
possible = (0.25, 0.35, 0.55, 0.65)
quite possible = (0.45, 0.55, 0.75, 0.85)
very possible = (0.65, 0.75, 1, 1)
almost sure = (0.95, 0.98, 1, 1)
sure = (1, 1, 1, 1)

The use of confidence factors combined with certainty factors associated to the fuzzy values increases the performance of the system as the experimental results show in the next section.

During the reasoning process, the inference engine will add new facts to the facts base and will associate to each fact its corresponding confidence factor, computed according to the previous rules (1÷3). In this process several paths could be followed (e.g. in the case of alternative solutions for a given problem). When the reasoning process stops, the system will choose the conclusion with the maximum confidence factor.

The rules from RB are classified into three categories: behavior, decision and control. In figure 4 it is represented the information flow (data flow & control flow) associated to the knowledge-based AEPS, with the highlight of the three categories of rules.



PC – Pollutants Concentrations
MD – Meteorological Data
IP&A – Information Processing and Analysis
BR – Behavior Rules
DR – Decision Rules
CR – Control Rules

Fig. 4. Information flow of the knowledge-based AEPS

Behavior rules are characterizing the status of the system in certain conditions. The general form of a behavior rule is the following:

RULE Behavior$_i$
IF <premise> **THEN**
        * the system is in state $s_i(t)$ - at time t;

Example of a generic behavior rule:

RULE B_{Pi}
IF $P_i \in [L_{inf}, L_{sup}]$ **THEN**
        * the status variable is within the suboptimality interval;

where P_i represents the parameter *i*, L_{inf} and L_{sup} are the limits of the preset interval.

The control rules are characterizing the changes of the system's state in certain conditions. The general form of a control rule is as follows:

RULE Control$_i$
IF <premise> **THEN**
        * the system passes in the next state $s_i(t+1)$;

Example of a control rule:

RULE Control$_{Ip}$
IF $T_{predicted}$ = much_higher **AND** $|I_P - L_{sup}| < \varepsilon$ **THEN**
        Risk_of_pollution = possible;

where $T_{predicted}$ is the predicted temperature value, the numerical value associated to the symbolic value *much_higher* is set by taking into account the period of the year and the maximum value of the temperature in that period, I_P is the predicted value for a pollution indicator, L_{sup} is the maximum acceptable value for that indicator.

The decision rules have the role of taking some decisions (i.e. actions) in order to have the monitored parameters inside their suboptimality interval. The general form of a decision rule is the following:

RULE $Decision_i$
IF <premise> **THEN**
* take a decision (action) in order to have the parameters inside their suboptimality interval;

Example of a generic decision rule:

RULE $Decision_{Ip}$
IF risk_of_pollution = possible **THEN**
* send a **warning**: area with potential atmospheric pollution risk!
* find a solution for a decrease of the pollution indicator;

Another category of rules is given by metarules that establish a strategy for applying other rules. An example of such rule is the following:

RULE 31:
IF season = summer **THEN**
* apply rule $Control_{Ip}$ with priority (or in general cases, rules that have in the premise a condition referring to an important increase of the temperature);

Examples of rules from RB:

RULE 57:
IF D_T > 7_days **AND** T >= 38 (C^0) **THEN**
$T_{predicted}$ = much_higher CNF 100;

RULE 22_5:
IF T = much_higher **AND** $|I_{P\text{-}MAC\text{-}sp} - L_{sup\text{-}MAC}| < 0.005$ **THEN**
* **very possible** exceed of the MAC_{SP} CNF 95;

where D_T is the duration of the predicted period with higher temperatures, and MAC_{SP} is the maximum admissible concentrations of suspended particulates (SP).

There are a lot of unpredictable events that may influence the atmospheric pollution degree and it is difficult to establish with certainty which are the causes of a decrease or of an increase of a pollutant indicator. Sometimes a decrease can have at the basis not an environmental protection management decision. It can simply be a temporary situation. For example, economical events (a plant is temporary closed or the production of a specific product is temporary reduced or stopped) or unpredictable meteorological events.

As a correct acquisition and structuring of the atmospheric environmental protection knowledge into a knowledge base is of great importance for the success of the knowledge-based AEPS, we have designed an atmospheric environmental protection ontology. We have developed a prototype of the ontology, $\mathbf{O_{AEP}}$, which includes also entities from the meteorological domain that have a direct influence on the atmospheric environment. The air pollutant substances (NH_3, NO_x, SO_2, fenol, H2S, CO, suspended particulates emissions etc) have associated their specific maximum admissible concentrations (MAC) whose values are taken from the air quality standards. All these information are entities from the ontology vocabulary (constants, predicates, etc). Another generic entity is the meteorological factors with possible values such as the temperature of the air (T), the speed of the wind (SW), the intensity of solar radiation (ISR) etc. The rules from the rules base describe relations between the entities of the ontology vocabulary. The prototype of the ontology, $\mathbf{O_{AEP}}$, has provided a clear and effective description of the knowledge base by structuring the concepts that are used in the domain of atmospheric environmental protection and related domains such as the meteorological domain.

4 Experimental Results

We have run several preliminary experiments of the knowledge-based AEPS on a set of test cases and a set of uniformly distributed sites on the area of Ploiesti, and we have compared the results with those obtained by human experts. The system analyses past data of the pollutant substances concentrations (measurements made in the last 10 years – 1990-1999) and predicts their values for a short term (for the period 2001-2002), by using a feed forward neural network. These values are correlated with the meteorological predictions and with the knowledge that exists in the knowledge base, and finally, the system generates the sites with possible atmospheric pollution risk, giving also some solutions.

We have driven the experiments on 7 representative sites (sites 1, 2, 5, 7, 8, 11, 12) from the area of Ploiesti, and the test cases included measurements of 5 pollutant substances concentrations. The periods of the analysis were 2000/2001 and 2001/2002 divided in the four seasons starting with autumn. We have determined the percentage of good detections of sites with atmospheric pollution risk.

Period	XII-II 01/02		III-V 2002		VI-VIII ‘02		IX-XI 2002	
Type	KBS	HE	KBS	HE	KBS	HE	KBS	HE
% GD	96%	97%	92%	95%	98%	96%	97%	98%

Type of experiment: KBS – Knowledge-based System, HE – Human Expert
%GD – percentage of good detections of sites with atmospheric pollution risk (rounded values)

Table 1. Experimental results 2001/2002

Table 1 presents the results obtained for the period 2001/2002. As we can see the knowledge-based AEPS had a good behavior. Still, the human expert (HE) gave better results (except summer 2001 and summer 2002), mainly because it uses some heuristics that are difficult to formalized, heuristics that proved to be good in the cases that were analyzed. So, as mentioned in section 3 some effort should be concentrated on a much proper expert knowledge acquisition and representation, extending also the developed ontology. In the summer of 2001 and in that of 2002 the meteorological forecasting system made good predictions and the rules that connect meteorological knowledge with atmospheric environmental knowledge were better than the expert knowledge used by the human expert. These rules were automatically generated by an inductive learning algorithm, which provided efficient rules.

SITE 11 – *Period*: *July 2002*
Predictions:
$T >= 38^0C$
$D_T > 7$ days (duration of the predicted period with higher temperatures)
SW = *weak*
Pollutants concentrations (measurements):
$SO_2 = 0.00861$ mg/m^3
($MAC_{SO2} = 0.25$ at 24h)
$NO_2 = 0.03257$ mg/m^3
($MAC_{NO2} = 0.10$ at 24h)
SP(suspended particulates) = 0.14693 mg/m^3
($MAC_{SP} = 0.15$ at 24h)
Fenol = 0.00823 mg/m^3
($MAC_{fenol} = 0.03$ at 24h)
$NH_3 = 0.05778$ mg/m^3
($MAC_{NH3} = 0.10$ at 24h)
Season = *summer*

Conclusion:
Warning: *very possible* exceed of the MAC_{SP}.
Decision & Control:
* Warning of the chemical plants that are responsible for the emissions of the suspended particulates in the air, in the area of site 11.
* Find preventive measures for reducing the emissions of suspended particulates (e.g. change of the chemical process technology, introducing new performant filters, temporary stop of the productive sector that produces the emissions until a long-term solution is found).

Fig.5. Example of *SBC_ProtMed* run

Figure 5 shows an example of the atmospheric environmental analysis and decision making at a specific location, site 11.

The experiments were made on sites that are in situ and as a future work we shall analyze the scalability of the system, i.e. the increased environmental loading due to an increase in the number of distributed components. This work will be done on a simulation that considers all sites on line.

The knowledge-based AEPS was implemented in C++Builder and the knowledge base of the *SBC_ProtMed* system was implemented in VP-Expert, an expert system generator.

5 Conclusions and Future Work

In this paper we have presented a knowledge-based system as an implementation of an atmospheric environmental protection system. The main role of the system is to provide expert assistance at the decision making level and also at the control level of the atmospheric protection management in the area of Ploiesti. At present, the developed KBS has the ability to detect with success the sites with atmospheric pollution risk and also can evaluate the possible solutions to reduce such risks. These are important benefits because the system could be used as an expert assistant in the atomspheric environmental management. The knowledge based system has the possibility to be connected with the commercial systems available to monitor pollution from the major air polluters in urban areas, such as MLU System, NILU System or the Japanese Horiba automotive analyzers.

As a future work, we should concentrate on building a better knowledge acquisition module and we shall include a full real time capability of the knowledge-based AEPS. Also, we shall extend the atmospheric environmental modeling and simulation subsystem with environmental simulators and other dispersion models for air pollution substances. Another direction for our future work is to include improving suggestions the system provides to the polluters in terms of what can be done to mitigate short-term pollution excess.

Taking into account the large amount of environmental data and the diversity of factors that influence the degree of atmospheric pollution in a specific area, the developed *SBC_ProtMed* system has a great utility in the context of improving the air quality in urban regions. Such a knowledge-based system extended with a full real time capability would be a powerful expert advice tool for the atmospheric environmental protection management.

Acknowledgement

The author is grateful to the Environmental Protection Agency Prahova who provided access to its measurements databases. The research reported in this paper was partially funded by the Romanian Education and Research Ministry and the National Council for Scientific Research - CNCSIS under the grant A_T no. 429/2003.

References

1. Canter L.W.: *Environmental Impact Assesment*, McGraw Hill (1997)
2. Dvorak D., Kuipers B.: Process Monitoring and Diagnosis, *IEEE Expert* 6(3):67-74, (1991)
3. Godo L., Lopez de Mantaras R., Sierra C., Verdaquer A.: MILORD: the Architecture and the Management of Linguistically Expressed Uncertainty, *International Journal of Intelligent Systems* 4:471-501, (1989)
4. Halker J.M. (eds):, *Proceedings 6th Symposium Computer Science for Environmental Protection,* Munich, Informatik-Fachberichte 296, Springer, Berlin (1991)

5. Kruse R., Schwecke E., Heinsohn J.: *Uncertainty and Vagueness in Knowledge Based Systems*, Springer, Berlin (1991)
6. Nakamori Y., Sawaragi Y.: Methodology and Systems for Environmental Decision Support, *Proceedings LSSTA Symposium*, Vol.1, London (1995) 1-11
7. Oprea M.: Some Ecological Phenomena Forecasting by Using an Artificial Neural Network, *Proceedings 16th IASTED International Conference in Applied Informatics*, Garmisch-Partenkirchen, Germany, (1998) 30-32
8. Oprea M.: Toward the Development of a Knowledge Based System for Environmental Protection, *Proceedings 8th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, Vol.II, Madrid, Spain, (2000) 1003-1008
9. Page B.: An Analysis of Environmental Expert Systems Applications, *Environmental Software* 5(4):177-198, (1990)
10. Riano D.: Learning Rules within the Framework of Environmental Sciences, *Proceedings ECAI'98 BESAI workshop*, Brighton, UK (1998)
11. Salski A.: Fuzzy Data Analysis in Ecological Research, *Proceedings 8th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, Vol.II, (2000) 997-1002