Επιχειρησιακή Έρευνα / Operational Research. An International Journal. Vol.5, No.1 (2005), pp.105-113

Decision Support System for a River Network Pollution Estimation Based on a Structural-Linguistic Data Model

A. Hachikyan[1], R. Scherer[2], M. Angelova[1], M. Lazarova[1]

[1]Technical University of Sofia, Department of Computer Science
8, Kliment Ohridski blv., 1000, Sofia, Bulgaria
gotchev@tu-sofia.bg, ahh@tu-sofia.bg, milaz@tu-sofia.bg
[2]University of Karlsruhe, Institute of Practical Mathematics
Kaiserstrasse 12, D – 76128, Karlsruhe, Germany
scherer@math.uni-karlsruhe.de

Abstract

On the base of a specially crated structural-linguistic model of a river network a point pollution detection decision support system (DSS) is developed. The data model completely describes the topology of the river and the presence of technical equipment along the river basin. The described system is able to find monitoring stations with pollution detected and estimates the potential polluting objects by means of several heuristic rules. The final result of the DSS analysis is an arranged list of polluting objects with the highest score for the most probable of them. The structural model of the river network is developed in the ArcInfo environment and Visual Basic is used for the construction of the DSS.

Keywords: structural-linguistic model, surface water purity

1. Assumptions

With the industry enlargement human activities cause alternation of the natural balance in many ecosystems. One typical example is the pollution of the water recourses (rivers, lakes and seas). This process has to be strictly controlled in order not to cause fatal changes. That is why the river basin water quality monitoring is an important and serious problem. The following initial conditions are assumed:

1. A structural-linguistic model for the river network is build. It completely describes the topology of the river and the presence of technical equipment (monitoring stations, polluting objects) along the river network. Geographic Information Systems evaluation environment ArcInfo is used for the model construction. A number of layers, describing all the necessary information, are created. Each one of them has a corresponding attribute table. The collection of all tables forms the data base of the model. The Struma river network is used as an example for the evaluation of the

structural-linguistic model, but it can be directly applied to any other river network. A full description of the model is given in [Gotchev (2003)].
2. The geographical position of the surface water quality monitoring stations is available.
3. A list of all monitored polluting parameters is defined.
4. The geographical position of all polluting objects along the river is determined and for each of them a list of all emitted polluting substances is known in advance.
5. The monitoring process is based on the analysis results derived from the surface water quality monitoring stations. Here is also included information for the date and the water amount used.
6. Only point pollution sources (objects with precise geographical position) are considered.
7. The defined by the official regulations for each river flow planed category is known in advance.

The final result of the DSS analysis is an arranged list of polluting objects with the highest score for the most probable of them. The system should be able to explain the suggested decisions.

2. Surface Water Quality Parameters

Water quality depends on the concentration of the polluting substances and the flowing water runoff. Depending on the river water usage there are four water categories:

- First category - best water quality, suitable for drinking, household and usage in the food industry
- Second category - water suitable for agriculture water-supply systems, for water sports and fishing
- Third category - water suitable for irrigation, industry water supply, etc.
- Fourth category – very polluted water, suitable only for water transport

Limit concentrations for all monitored polluting substances are defined by the official regulations. Physical and chemical analyses of the water are made in the surface water quality monitoring stations. The water samples are observed in laboratories for the presence of chemicals, heavy metals and organic pollution. The condition of all main water indices such as dissolved oxygen, (acid level) PH, oxidation, etc. are also estimated [PHARE (1998)].

In the described DSS all 26 monitoring stations located on the main river flow and the first order river flows of Struma are examined and the condition of 14 polluting substances is monitored.

3. Heuristics

When a monitoring station with pollution detected is discovered the system decides which are the probable polluting objects. All of them are situated upstream the monitoring station. In order to make the suggestions the system uses two heuristic rules, which add to the total pollutant score certain amount of points (positive real numbers).

3.1. Distance

The degree of influence of each polluting object over the water quality observed in the examined monitoring station depends of the geographical distance between the two objects because of the self-purifying ability of the river. The following production rules describe this effect [Hachikyan (2002)]:

i) if (distance<=20 km) then points=3
ii) if (distance>20km And distance<=50km) then points=2
iii) if (distance>50km) then points=1

3.2. Intersection between the detected polluting substances exceeding the a priori defined limits and the pollutants emitted in the river by each polluting object

The polluting substances for which higher concentration is detected form a subset of the set of all monitored polluting parameters. For each pollutant a list of all emitted substances is defined. The emitted substances also form a subset of the set of all monitored polluting parameters. The intersection between the two subsets is evaluated using the following rule [Hachikyan (2002)]:

i) if (match=all indices) then points=3
ii) if (match=more than half of the indices or equal to the half) then points=2
iii) if (match=less than half of the indices) then points=1
iv) if (no match) then points=0

The DSS has a forward chaining (data-driven) reasoning. When the if (condition) part of the rule is fulfilled by the facts about the current situation, the rule is fired and its then (action) part is executed [Negnevitsky (2002)]. The final evaluation for each pollutant is the sum of the points appointed for both rules.

4. Algorithm Development

For simplicity the algorithm is presented as a description of the applied steps. The main problem is divided in two parts, which are treated by the system in a consequence. The expert system derives all the necessary information from the data base of the structural-linguistic model.

4.1 Specification of the monitoring stations with pollution detected for the given period of time.

The research is made for a selected month of the year due to the season distribution of the water amounts. One of the parameters of each monitoring station is the planned water category. The pollution is detected when the actual water category is worse than the planned category.

The algorithm consists of the following steps:

1. Extraction of the data for the given month and the currently examined monitoring station from all water analysis information.
2. Calculation of the monthly average load of each polluting substance for the given month k.

$$P_j^k(mg/s) = \frac{1000}{N} . \sum_j Q_j^k . C_{ji}^k \tag{1}$$

$i = 1, 2, \ldots, 14$

i – polluting substance

Q_j^k - water amount taken from the river during the $j^{\text{-th}}$ measurement (m^3/s) for the given month k

N – number of the measurements accomplished for month k

3. Calculation of the monthly average concentration of the $i^{\text{-th}}$ polluting substance given the monthly minimal water amount flowing in the river Q_{95}^k

$$C_i^{k,95}(mg/l) = \frac{P_i^k}{Q_{95}^k.1000} \tag{2}$$

4. Determination of the real water category for the currently examined monitoring station for each polluting substance.
The real categories are specified in accordance with the limit concentrations defined with official regulations.
5. Determination of the final real water category for the currently examined monitoring station.
If the real category for one of the polluting parameters in the current monitoring station is worse than the planned one, the final water category is equal to the real category.
6. Recording the calculation results for all polluting substances in the current monitoring station.
7. Comparison between the real (actual) and the planned category for the currently examined monitoring station.

The above described sequence of actions is performed for all 26 monitoring stations.

4.2. Search for the relevant polluting objects and assigning points to each one of them.

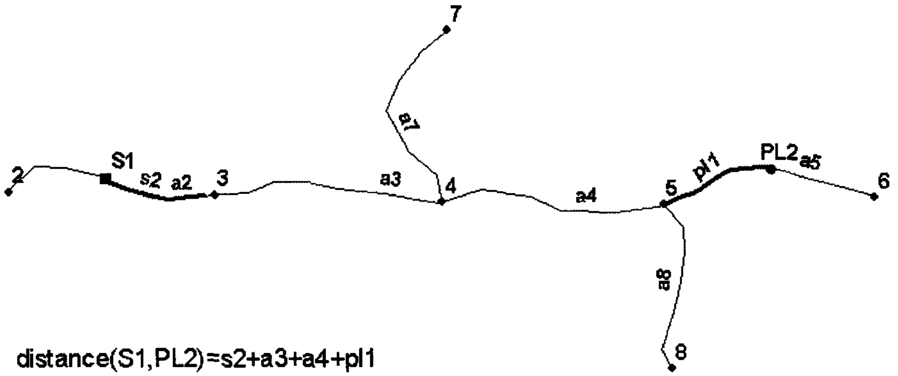
A structural model of river network of a basin is represented as oriented graph. The nodes depict the structural points and the oriented arcs are the segments of the main river and its flows to a certain level. The structural points are the points where a flow reaches the main river or one flow reaches another, the points of branching. This structural model is described by means of a formal linguistic context dependant system. The structural-linguistic model can be regarded as an element of GIS. The presence of technical equipment along the river (monitoring stations, polluting objects) is reflected by as adding technical nodes and technical arcs to the model. The technical nodes show the geographical position of the equipment. The technical arcs are results of the separation of the arcs by the presence of technical nodes. The bigger objects as lakes, dams and isles are described as polygons (closed sequence of arcs). The described presentation of the river network is successfully used in the expert evaluation of the system.

The algorithm consists of the following steps:

1. Selection of one monitoring station with pollution detected.
2. Determination of the arc together with its beginning and ending structural node, which contains the selected monitoring station.
3. Determination of the sequence of arcs and structural nodes upstream the river located between the monitoring station and the spring of the main river or the highest order river flow.
4. Detection of all polluting objects located on the main river or highest river flow.
5. Calculation of the distance between the selected monitoring station and the detected pollutants.

   The above described operation is accomplished by following the road of arcs and technical segments between the two points and summing their shape lengths (Figure 1).

6. The existence of lower river flows (right or left) is checked for each structural node from the followed main river or highest order river flow. If there exist such flows then the connecting structural nodes (4, 5), connecting arcs (a7, a8) and end nodes of the arcs (7, 8) for each one of them are determined.
7. If a polluting object is located on any of these arcs, the distance between the object and the monitoring station is calculated using the road of arcs presentation.
8. The described actions (from step 2 to step 7) are recursively repeated for all detected lower river flows. The starting node is the end node of the connecting arc (6, 7). The exiting recursion condition is the reach of the end of the flow (no next structural nodes).



ai – arc, i=1, 2, …, M
M – total number of arcs
s2, pl1 – technical segment
S1- monitoring station
PL2 – polluting object

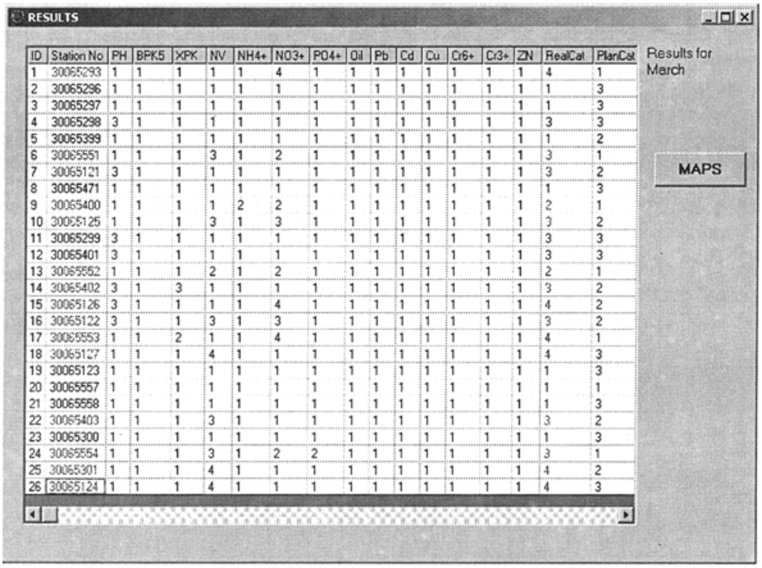
Figure 1. *Road of arcs*

9. Points are assigned to each probable pollutant according to the calculated distance (heuristic rule 1).
10. Points are assigned to each probable pollutant according to the results from the intersection between the subset of the polluting parameters with high concentration detected and the subset of the polluting substances emitted by each polluting object (heuristic rule 2).
11. Sum of the points assigned to each pollutant in the previous two steps (step 9 and step 10).
12. Arrange the list of the probable pollutants in descending order of the final sum, calculated in step 11.

All current and final results are recorded by the system, which makes possible the explanation of the expert evaluation.

5. Results of the System Functioning

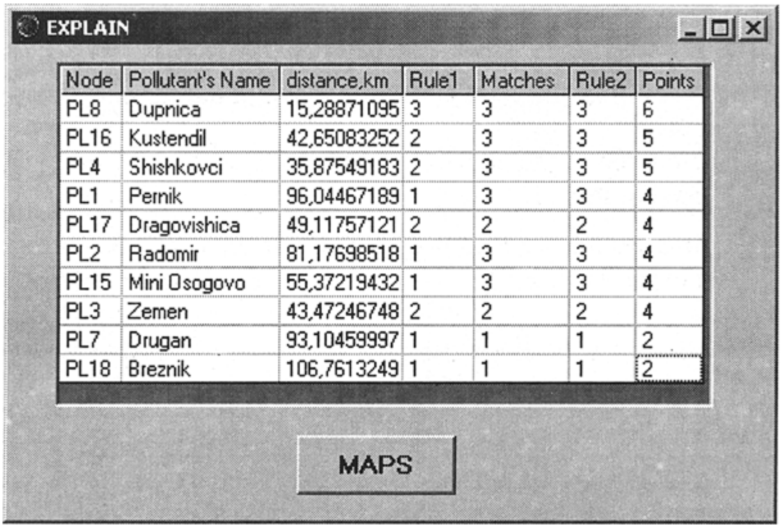
The DSS was tested with data acquired by real measurements received by the surface water quality monitoring stations for the Struma river basin. After the month selection the real water category for each monitoring station is calculated and it is compared with the planned category for the river water at the same station (Figure 2).



ID	Station No	PH	BPK5	XPK	NV	NH4+	NO3+	PO4+	Oil	Pb	Cd	Cu	Cr6+	Cr3+	ZN	RealCat	PlanCat
1	30065293	1	1	1	1	1	4	1	1	1	1	1	1	1	1	4	1
2	30065296	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
3	30065297	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
4	30065298	3	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3
5	30065399	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
6	30065551	1	1	1	3	1	2	1	1	1	1	1	1	1	1	3	1
7	30065121	3	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2
8	30065471	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
9	30065400	1	1	1	1	2	2	1	1	1	1	1	1	1	1	2	1
10	30065125	1	1	1	3	1	3	1	1	1	1	1	1	1	1	3	2
11	30065299	3	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3
12	30065401	3	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3
13	30065552	1	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1
14	30065402	3	1	3	1	1	1	1	1	1	1	1	1	1	1	3	2
15	30065126	3	1	1	1	1	4	1	1	1	1	1	1	1	1	4	2
16	30065122	3	1	1	3	1	3	1	1	1	1	1	1	1	1	3	2
17	30065553	1	1	2	1	1	4	1	1	1	1	1	1	1	1	4	1
18	30065127	1	1	1	4	1	1	1	1	1	1	1	1	1	1	4	3
19	30065123	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
20	30065557	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	30065558	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
22	30065403	1	1	1	3	1	1	1	1	1	1	1	1	1	1	3	2
23	30065300	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
24	30065554	1	1	1	3	1	2	2	1	1	1	1	1	1	1	3	1
25	30065301	1	1	1	4	1	1	1	1	1	1	1	1	1	1	4	2
26	30065124	1	1	1	4	1	1	1	1	1	1	1	1	1	1	4	3

Figure 2. *Distribution of the water categories and the polluting indices for all monitoring stations*

After that the arranged list of probable polluting objects for a chosen monitoring station is created and the system informs the user what are the results from the calculations for each pollutant (Figure 3).



Node	Pollutant's Name	distance,km	Rule1	Matches	Rule2	Points
PL8	Dupnica	15,28871095	3	3	3	6
PL16	Kustendil	42,65083252	2	3	3	5
PL4	Shishkovci	35,87549183	2	3	3	5
PL1	Pernik	96,04467189	1	3	3	4
PL17	Dragovishica	49,11757121	2	2	2	4
PL2	Radomir	81,17698518	1	3	3	4
PL15	Mini Osogovo	55,37219432	1	3	3	4
PL3	Zemen	43,47246748	2	2	2	4
PL7	Drugan	93,10459997	1	1	1	2
PL18	Breznik	106,7613249	1	1	1	2

Figure 3. *Final results with an explanation for each polluting object*

In the end the final results are depicted on a map of the river basin. This helps the user better to estimate the credibility of the decisions made by the system (Figure 4).

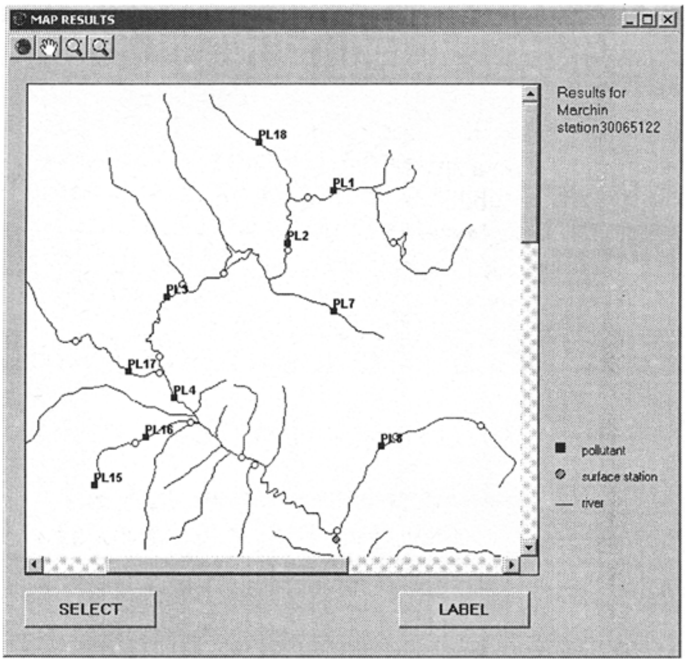


Figure 4. The final results, depicted on a map

6. Conclusion

The described DSS can be used for a quick general evaluation of the river water quality in the entire river basin. The arranged list with probable polluting objects is a good starting point when seeking the reasons for water pollution. It can ease the work of people who are not specialists as well as this of experts in the domain. The use of structural-linguistic model as a base for the river network topology exploring leads to the development of an appropriate algorithm. The calculation of the distance between two river points is performed as simple sum of shape lengths of the arcs and technical segments located between the points. The presence of the technical equipment along the river (monitoring stations, polluting objects) as technical nodes in the structural model significantly assists the process of finding the polluting objects in all parts of the river. This is of further help when solving the mentioned problem. The used structural-linguistic model of the Struma river network includes only the main river and the first order river flows, but it can be easily used for any order river flows included. To do this the maps and attribute tables have only to be completed with all necessary additional data. The developed recursive algorithm of the system implementation is able to process river flows of any order. The described expert evaluation can easily be applied to any other river basin having a structural-linguistic description developed. All obtained results in the described evaluation are used in INCO-Copernicus project "WATERMAN" Contact Number IC15 CT98 0107.

References

Gotchev, G., Angelova, M. (2003), "A Structural-linguistic Model and Data Base Description and Presentation of a River Network of a River Basin in the Geographic Information Systems", Automatics and Informatics, 7, pp.23-27.

Hachikyan, A., Scherer, R., Bojchev, D., Gotchev, G. (2002), "Expert System for Ecological Monitoring of a River Basin", Applications of Mathematics in Engineering and Economics, 27, Heron Press, Sofia.

Negnevitsky, M. (2002). Artificial Intelligence A Guide to Intelligent Systems. Addison-Wesley.

PHARE official report (1998). A Review of the trans-border river pollution in South Bulgaria, vol.1 and vol.3.