RRB 89-39

This manuscript has been prepared for
the Proceedings of
IASTED International Conference on Expert Systems
and the contents are subject to change.

This copy is to provide information
prior to publication.

ACID RAIN TO MINE EFFLUENT: EXPERT SYSTEMS APPROACH TO ENVIRONMENTAL MODELLING

by
D.C.L. Lam, D.A. Swayne*, I. Wong,
A.S. Fraser, Y. Ando*, J. Storey*,
and J. Kerby*

National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario, L7R 4A6

*Department of Computing and
Information Sciences
University of Guelph
Guelph, Ontario

May 1989
NWRI Contribution #89-110

MANAGEMENT PERSPECTIVE

The RAISON system was originally designed for the regional analysis of the impact of acidic deposition for use on a microcomputer. Its unique feature includes a user friendly interface of database, map, and spreadsheet. This interface has facilitated the analysis of air, soil and water data and the presentation of watershed acidification model results. The success of using such a system has led to new applications. This report shows how the system is adapted from the acid rain problem to the mine effluent problem. It illustrates the flexibility and the general applicability of the system.

PERSPECTIVE-GESTION

Le système RAISON a été conçu à l'origine pour la réalisaton d'analyses régionales de l'impact des retombées acides à l'aide d'un micro-ordinateur. L'existence d'une interface conviviale pour la base de données, les cartes et le tableur caractérisent ce système. Cette interface a facilité l'analyse de données relatives à l'air, au sol et aux eaux et a permis de modéliser l'acidification d'un bassin hydrographique. Le succès obtenu avec ce système a permis de nouvelles applications. Ce rapport montre comment ce système, d'abord conçu pour le problème des pluies acides, peut être adapté aux problèmes des effluents miniers. Il illustre la flexibilité et la grande diversité des applications possibles de ce système.

RÉSUMÉ

Un prototype de système expert, le système RAISON, a été appliqué à deux problèmes environnementaux. Les résultats préliminaires montrent que, en raison de la complexité et de la multidisciplinarité de la base de connaissances et de la diversité des données dans le domaine des sciences de l'environnement, il faudrait améliorer les liens entre les sous-systèmes de soutien (cartes, base de données, tableur etc.) avant de travailler à la mise au point d'une composante d'inférence interactive.

ACID RAIN TO MINE EFFLUENT: EXPERT SYSTEMS APPROACH TO ENVIRONMENTAL MODELLING

D.C.L. Lam, D.A. Swayne, I. Wong, A.S. Fraser, Y. Ando*,*
J. Storey and J. Kerby**

National Water Research Institute, Burlington, Ont. Canada
*Dept. of Computing and Information Sciences, U. of Guelph,
Guelph, Ont., Canada

ABSTRACT

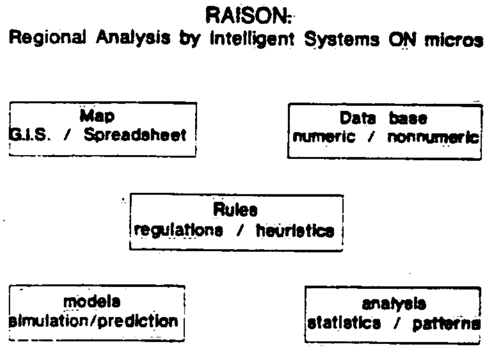
A prototype expert system, RAISON, has been applied to two environmental problems. Preliminary results show that, because of the complex, multidisciplinary knowledge base and the diversified data requirements in environmental sciences, efforts spent on linking up the supporting subsystems such as map, database and spreadsheet are required, prior to the development of an interactive inference component.

KEYWORDS: expert systems, acid rain, mine effluent, environmental science, models

INTRODUCTION

The application of the expert systems approach to environmental problems has emerged only recently. There are a number of potential application areas, e.g. regional assessment and emergency spill management. There are excellent quantitative and descriptive information for such applications. In practice, however, the development of environmental expert systems is difficult, because environmental science is highly multidisciplinary in nature and the type of information is very diversified. A really working expert system for most environmental problems needs an effective database subsystem, a geographical information system (G.I.S.), and an inference capability. The purpose of this paper is to present a prototype expert system that incorporates all these features and to provide examples of application in two important environmental issues, namely acid rain and mine effluent.

Figure 1. RAISON subsystems.



THE RAISON SYSTEM

We have developed a prototype expert system [1,2], acronymed RAISON, for Regional Analysis by Intelligent Systems ON a microcomputer (e.g., IBM/PC AT). This system is developed basically to meet the demand of environmental applications in which the knowledge base involves physical, chemical and biological disciplines and the data cover the air, soil and water regimes. To facilitate the flow of information among the disciplines and regimes freely and interactively, RAISON is designed with five subsystems, all interlinked (Figure 1): map, database, inference rules, models and analysis. It is the interlinking capability that makes RAISON different from existing individual subsystems such as some G.I.S. While these individual G.I.S. subsystems can be linked to a database system, they may not be able to connect to an expert system shell or complicated mathematical models. In RAISON, with all five components interlinked, the application from one problem to another can be achieved readily.

The Map (G.I.S.) Subsystem

In RAISON, the map subsystem helps in identifying inputs and in presenting the outputs of the expert system. Many environmental problems requires the map for data presentation (sometimes the data themselves are stored as maps). For this purpose, the maps are based on raster images to enable easy access to the database via the map. In RAISON, different levels of spatial detail can be revealed by zooming in or out. By displaying the stations at the appropriate level in the map, one can obtain access to the database (i.e. time series) at any of these stations. The data can be classified, analysed and modelled, and the results can be plotted as graphs, or backcoloured on the station, region or anywhere on the map. The data can be retrieved according to the overlayed zones of two or more maps. The ability in RAISON to draw contours and to define influence zones expedites the application of knowledge rules that require such overlay map analysis.

The Database

The RAISON database is one that allows different types of information, e.g. numeric data and descriptive text, to be input into the system. The numeric data are entered with a user defined layout that conform to the format of the data. It can be linked with a location such as a station or a district on the map subsystem, so that access to the database can be gained from the map subsystem. Once accessed, the data can be transmitted to a spreadsheet facility in RAISON for simple row and column computations. For more lengthy computations, e.g. running a model for all the stations enclosed within a given zone, a computer program can be written either in the C Language, or in a more simple language called the RAISON Programming Language (RPL), specially designed for this purpose. For the textual records, a narrative format is used so that search and match for a word or a phrase can be performed. In the development of the interface between the logical inference subsystem and the database, a prototype natural language translator program has been developed to convert recommendations produced by expert system shells and written in natural language into an RPL program. This interface RPL program will execute the recommendations and process both the numeric data and the descriptive records.

The Inference Capability

The inference capability is the focus of the whole expert system. It is where knowledge from the domain experts is integrated and logically presented. It is our opinion that there exist many readily available expert system shells that are useful for environmental applications. We do not need a complicated and elaborate shell; we only need a simple and functional one, at least for the present stage of development. As many of our applications require the use of a logical tree structure to organize our thoughts, we tend to favour the use of those shells (e.g., First Class, Nexpert, etc.) that provide such a capability effectively. We feel that to be able to deliver the results, not only is the choice of the shells important, but also we need the ability to translate the recommendations from the shell into an executable form with linkage to the map, the database and the analysis components. This is particularly true when the rules have to be operated in both the forward and the backward directions iteratively. We cannot afford to switch manually from one subsystem to another in processing the rules against a large and diversified database.

Simulation Models

In the area of environmental sciences, simulation models are commonly used as a research tool for analysing the problem and for making predictions [2]. Since the environmental problems are complex, to build an all-encompassing model that incorporates all possible processes is difficult. Most models, therefore, tend to be specific, limited by the model assumptions and data availability. In the expert system approach, we use the concept of having more than one model resident in the system so that selection of models can be made according to some heuristic rules such as matching the assumptions to the input data. This step assures the models are the most appropriate for the given assumption and data. However, the difficulty of the expert system approach is to define the rules. In RAISON, an iterative approach is used to refine the rules with the assistance of the interlinked map, database, inference and analysis components.

Analysis

The analysis subsystem in RAISON consists of statistical functions to provide statistical information (e.g., mean, median, regression and clustering analysis) on the spreadsheet. These functions can be used to compare model results with data, particularly for the validation of the expert system results with observed data. Special graphs can be generated and presented on the map subsystem (e.g. backcolouring the statistical information on the station or regional level), so that the regional differences can be detected effectively.

ACID RAIN APPLICATION

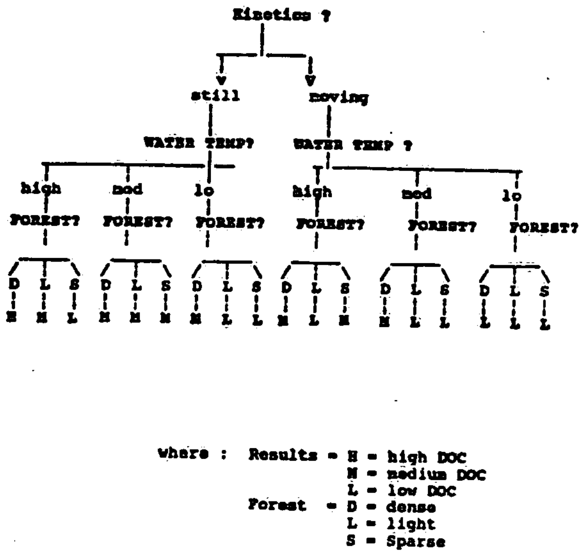
Acid rain is a major environmental issue in Northern America, Europe and other regions. The problem arises from the emission of sulphur dioxide and nitrogen oxide from industrial and other sources. These chemicals are transported in the atmosphere over large distances and are deposited, both in the dry and wet forms, onto the land surface. Most of them fall onto terrestrial regimes such as forestry canopy and different soil layers before entering streams and lakes. During this hydrological passage, particularly through the soil layers, the acids may be neutralized by the base cations present in the medium. The surface water quality conditions may be also affected by the in-lake or in-stream processes. One of the main objectives of the acid rain study is to predict the water quality condition based on a number of management scenarios of emission control. To achieve this goal, information from air, land and water has to be integrated and the predictions have to be made by selecting the most appropriate models according to a set of heuristic rules and the overlayed map system [2].

A preliminary attempt at applying the RAISON system to the acid rain problem has been reported in [2] and [3]. In this paper, we present some new features in the RAISON application.

The Knowledge Base

In the preliminary attempt [2], the knowledge rules are based on simple partition values of the Acid Neutralizing Capacity (ANC). Currently more indepth rule bases are constructed with existing expert system shell. An example is the rules for the incorporation of the influence of organic acidity associated with dissolved organic carbon (DOC) matters produced in bogs and forests. Figure 2 shows a possible tree structure, as constructed by the expert system shell called "First Class", relating the condition of DOC to water movements, water temperature and forest coverage. These rules are based on knowledge provided by domain experts in this area of study. The organic acidity is important in acting as a buffer against the inorganic acidity caused by the acid rain [4]. Thus, for relatively still water systems such as bogs and wetland, the production of DOC is known to be higher than streams or lakes with moving water. In general, warmer water temperature and greater abundance in coniferous trees are conducive to higher DOC. The results (Figure 2) as recommended by the "First Class" expert system are based on the logical combination of these conditions. These recommended results are fuzzy as the DOC concentration is separated into only three broad classes (high, medium and low) and the outcome could be different if more rules and factors are introduced. Nevertheless, based on the recommended classification, we can select an appropriate organic acid submodel (e.g. [4] for the high and medium classes) to be used in conjunction with the inorganic acidity model [2].

Figure 2. Example of a tree structure used for relating DOC concentration to water flow, temperature and forest.



The Inference Mechanism in RAISON

Other sets of knowledge rules are also designed, e.g. for the selection of snowmelt submodels based on episodic event probability and seasonal patterns. These rules are all implemented in the "First Class" expert system and the recommended results are translated within the RAISON system into RPL computer programs. These programs then constitute the core program that can actually execute the recommended rules in the forward or backward directions. In the forward direction, i.e. forward chaining, the data at each station are used to predict the ANC concentration for a given emission scenario. These predictions are made on the assumption that the recommended classification given by the rules (e.g. Figure 2) is acceptable. In the backward direction, or backward chaining, the assumption is considered to be tentative. For example, the predicted ANC concentration could be unacceptable if the model choice was not appropriate because of the fuzziness (i.e. several possible outcomes) in the classification. If so, the rules are applied in reverse direction to find out what possible input conditions could have led to a desirable ANC value. A new model is then selected.

Figure 3. Example of a layout program in RAISON.

```
Sample Problem Layout
 For [current_loading] AND [reduced_loading] do:
   For each station:
     if no-data then Generate_Data;
     calculate models;
          consult: model 1 rules
                   model 2 rules
                   ...
                   model n rules;
          if model choice unique then accept
          else use choice mechanism and mark;
   For each station:
          if spring then apply correction;
   For each station:
          if current_loading
          then calculate relative error;
   If reduced_loading AND anomaly then
            backchain to select new model.
```

Figure 3 summarizes the logical inference program layout in the RAISON system that considers two loading scenarios (current and reduced) and selection of models are made from several sets of rules. For unique answers, the forward chaining is used; for fuzzy answers, all the alternate choices are marked to prepare for subsequent backward chaining. Seasonal correction is made and statistical analyses are performed for comparing the two scenarios. In those cases where anomaly is detected, the rules are applied in reverse direction to find out what alternate models would have led to a more acceptable answer, consistent in both scenarios.

The Database and Analysis

The data base used for the present study is much greater than that used in [2]. In the preliminary analysis, only 364 stations for water chemistry for southern Quebec [2] are used. In the present study, a total of 5218 stations for lake water quality are used for Eastern Canada (Figure 4). Other data sets include those of deposition scenarios and soil sensitivity which are represented by contoured maps. The logical program as depicted in Figure 3 is applied to each of the 5218 stations. While the predicted results can be summarized by tertiary watersheds as in the case of [2], a more compact analysis is now based on aggregates of stations. For example, Figure 5 shows, under current emission scenario, the predicted aggregate ANC concentration (ueq/l) presented as statistical box plots (the top and bottom of the box represent the 25 and 75 percentiles and the middle bar in the box indicates the median). The outlaying values (10 and 90 percentiles) are denoted by the whiskers. This representation facilitates the comparison of the distribution of the ANC concentration of the aggregates. For example, using the 100 ueq/l as a standard, one can easily see which aggregate is in fair condition (e.g. AG12, a buffered region with high acidic deposition) and which one is in poor condition (e.g. AG1, a high organic acid area, cf. Figure 2, with relatively low acidic deposition).

Figure 4. Lake water quality stations.

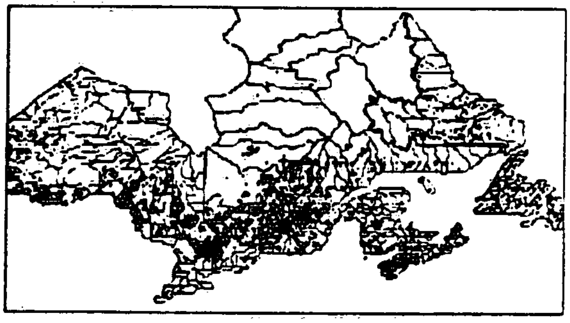
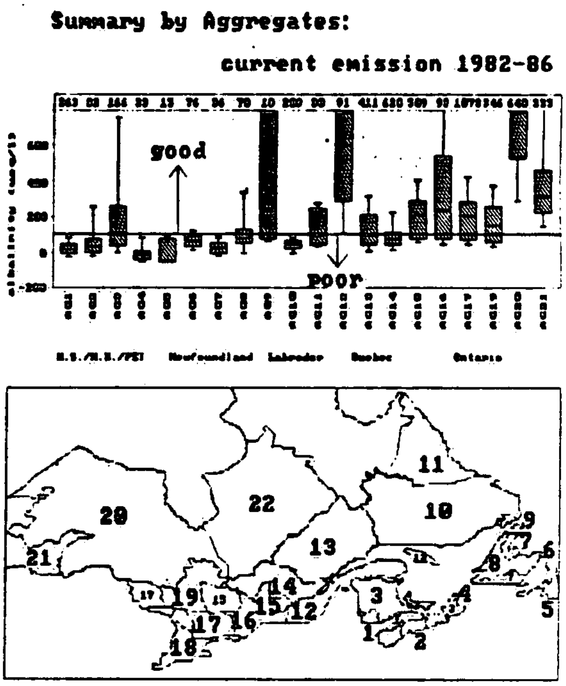


Figure 5. Predicted ANC concentration summarized for aggregates shown.



Thus, as demonstrated in these figures, the development of an expert system for the analysis of the impact of acidic deposition requires the construction of a well-linked system. An effective interface among the knowledge base, the rule construction, the database, the map system and the analytical and modelling capability is a necessity.

MINE EFFLUENT

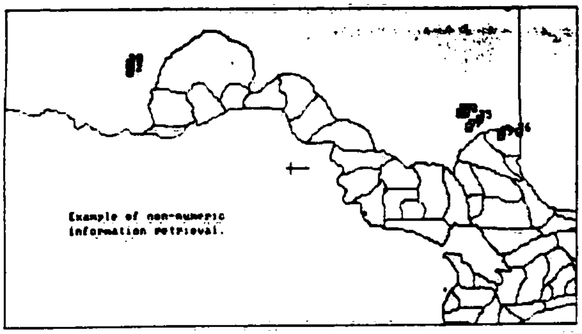
Following the success of the preliminary application of the RAISON prototype to the acid rain problem, we have started applying RAISON to another environmental issue, namely the impact of mine effluent on aquatic ecosystem. The modification to RAISON is trivial (e.g. remapping the region, etc.) and the mine effluent and toxicity data have been successfully transferred into the RAISON database. The main objective of this study is to allow domain experts to restucture the system so that the regional assessment for different management scenarios can be carried out. The main concern is how the chemical effluent are transported and transformed in streams and lakes and what level of effluent discharge could adversely affect the ecosystem, particularly in the case of multiple sources. Answers to these questions can affect future design of the monitoring network, i.e. where should the mine effluent be sampled to offer the most information. While the system is still being developed, some preliminary results [5] have been obtained that could further demonstrate RAISON's capability and flexibility.

Database

Much of the data are provided by the Ontario Ministry of Environment, our collaborator in this study. Again, both numeric data and descriptive records are entered to the RAISON system. The numeric data are mainly chemical effluent concentration measured in streams and lakes. The descriptive records pertain mainly to the mine operation and the location and geology of the area. Retrieval of the data is effective, as in the case of acid rain study. For example, Figure 6 shows all the mine locations where the word "zinc" is used in the textual database. The operation involves only a few commands and the search and display of the results on the map subsystem is fast. Similar effective access to the database and interactions with the map system will facilitate the processing of knowledge rules for this study.

Figure 6. Retrieval of textual data.

Locations containing the word "zinc"



Analysis

The next step we test is the statistical capability in RAISON. Often, after the data are successfully entered and the retrieved, simple statistical methods are performed on the data to derive empirical relationships. The relationships can be useful for building the rules for model selection. For example, the spreadsheet computation can be made, relating the flow data and nickel concentration at different stations. Screening of the data according to zones defined by overlay analysis or according to time windows is also possible. These analytical results provide valuable information about the spatial and temporal patterns in the data.

The Map Subsystem

For mine effluent analysis, a detail map of the surrounding environment is needed. For example, a detailed map for the mines near the city of Sudbury is shown in Figure 7. This map can be accessed from another map with a larger scale, e.g. Figure 6, if such details are warranted. The data and information on the hydrology, the land contours, the road bridges and the city blocks could lead to an accurate simulation of the river flows and the transport of chemicals along the river.

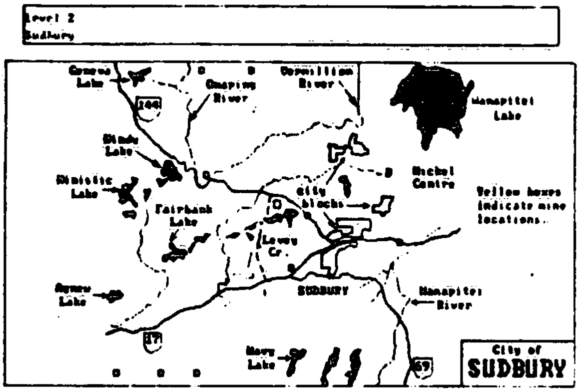


Figure 7. Detailed map for mines near Sudbury, Ontario.

Other Subsystems

We have already had some preliminary success of applying multi-variable transport models for the river system for a time-dependent simulation of the transport and pathways of the mine effluent. We are currently gathering the knowledge base from the domain experts to determine the rules for the model selection and the classification of regions according to the impact (e.g. toxicity) to the aquatic ecosystem (e.g. fish). Some of the rules are actually water quality objectives or regulatory requirements. Our approach is to build up the database, the map support, and the analytical and modelling capability first, before we can interact with the inference mechanisms. For the mine effluent study, we are in such a preparatory stage.

CONCLUSIONS

We have demonstrated the use of an expert system prototype for two environmental applications. The importance for a working system lies upon a well-linked system, with free and effective flow of information among the database, the analytical, the map, the modelling and the inference subsystems. The implementation is achieved in stages. Our experience shows, at least for the environmental problems we deal with, that the supporting subsystems must be set up and made operational first, before any of the inference component can be implemented.

ACKOWLEDGEMENTS

The study on mine effluent is supported by the Ontario Ministry of Environment (MOE) under the MISA Project. The authors are thankful to the assistance by J. Kinkead, Z. Novak, G. Bowen and S. Klose of the MOE.

REFERENCES

1. D.A. Swayne and A.S. Fraser, "Development of an expert system/intelligent interface for acid rain analysis", J. Microcomputers in Civil Engineering, 1, 1986, pp. 19-22.
2. D.C.L. Lam, D.A. Swayne, J. Storey, A.S. Fraser and I. Wong, "Regional Analysis of Watershed Acidification using the expert systems approach", Environ. Software, 3(3), 1988, pp. 127-134.
3. D.A. Swayne, D.C.L. Lam, A.S. Fraser and J. Storey, "Regional Acidification Models using the expert systems approach", Proc. IASTED Int. Conf. Expert Systems Theory and Applications, Geneva, Switzerland, 1988, pp. 19-22.
4. D.C.L. Lam, A.G. Bobba, R.A. Bourbonniere, G. Howell and M.E. Thompson, "Modelling organic and inorganic acidity in Nova Scotia Rivers", Water, Air, & Soil Pollution, 1989, (accepted).
5. I. Wong, D.C.L. Lam, G. Bowen and Z. Novak, "Impact assessment of mine effluent on aquatic ecosystems using RAISON-micro system", 1989, (in preparation).