Decision Support Systems for Water Resources Problems

David Lam

National Water Research Institute, Environment Canada, P.O. Box 5050, Burlington, Ontario, CANADA L7R 4A6
(Email:David.Lam@cciw.ca)

Abstract Environmental problems are complex. Their solutions require information on physical, chemical and biological processes, as well as considerations for socio-economic factors. A new approach is proposed for the development of decision support systems for environmental water resources problems by integrating the quantitative and qualitative knowledge on these scientific processes and socio-economic constraints. It was found that knowledge-based systems tools such as expert systems and neural network facilitated greatly the application of numerical models and data by using the associated descriptive knowledge and enabled better understanding between scientists and managers. Also, as a result of the recent paradigm shift on information technology towards microcomputers and the Internet, it is now possible to customarize the decision support system for a given problem by assembling the necessary knowledge-based tools and by accessing a variety of databases including real-time data, Geographical Information System (GIS) maps. This generic tool-kit approach of constructing decision support systems helps highlight the visualization of the spatial and temporal dynamics in the combined knowledge-bases. In this paper, several case studies including watershed management, acid rain and other water resources application were used to illustrate the approach.

1. INTRODUCTION

Environmental problems such as water resources allocation, irrigation, sediment erosion, agricultural production, toxic waste, industrial effluent, acid rain and climate change are complex issues. To understand them, we require the knowledge on many of the physical, chemical, geochemical, biological, benthic and ecological processes. To solve them, we need to consider management options such as resources diversion, pollution abatement and emission reduction, as well as socio-economic factors, in order to arrive at a solution with minimal cost for the policy implementation while achieving maximum benefit for the conservation and protection of the resources at risk. However, there are difficulties in bringing together data for physical, biological and other scientific disciplines, because generally they were sampled for the understanding of the individual processes, not as an integrative study. Often, there are problems of matching data because of the different time and spatial scales used. Similarly, when models from different disciplines are connected, the scale problem is one difficulty and the matching of modelling assumptions and applicability is another. To integrate these data and models, one must go beyond the quantitative knowledge and make use of the qualitative or descriptive knowledge. For example, based on the descriptive information or so-called meta-data files describing conditions or circumstances under which the data were sampled, we can set up screening rules for the inclusion or exclusion of one data set to be integrated with another data set. Another example is that, based on model assumptions or from the experience of the modellers, we can set up rules for choosing the most appropriate one from a selection of models for a given set of input data and conditions (Lam et al., 1989). The use of descriptive knowledge is particularly helpful when the disciplines (e.g. physics and economics) have limited basis for quantitative communication, or if the complexity of the sciences needs to be explained in simple terms for managers. In this paper, we propose a new framework by which the quantitative information from data and models can be combined with the qualitative knowledge to facilitate decision making.

2. A NEW FRAMEWORK

2.1 What is a Decision Support System?

A decision support system is a computer software that helps decision makers, managers and advisors find relevant information, accurate summaries, intelligent advice, optimal solutions, risk analysis and scenario gaming during the decision process. The decision support system itself does not make decision. It is designed to perform a supporting role to help managers and stakeholders make decision. As a simple example, a scientist may use a semi-empirical, cause-effect relationship between a chemical pollutant and an ecological indicator quite effectively for decision support purposes. For more complicated cases, many modelling studies were adapted to perform a decision support role by using the models to answer 'what-if' questions. Recently, because of the ability to visualize spatial details, geographical information systems (GIS) have attracted many scientists and managers to its potential as a decision support tool. However, for applications with diverse disciplines such as environmental water resources problems, GIS or mathematical models alone do not constitute a decision support system; they need a holistic sense of the problem being solved and therefore intelligent systems. They need descriptive knowledge.

2.2 The knowledge-based System Approach

One way to see the holistic picture of the problem being solved in the context of decision support is to look at a decision support system as one with input and output, using a number of tools that make use of the input to produce the output (Figure 1). On the one hand, environmental decisions are often based upon the interpretation, integration, presentation and classification of data and information, as well as the results of scenario testing, cost-benefit and risk analysis, to arrive at recommendations for the decision makers and stakeholders. On the other hand, the input at our disposal are in different types and forms: numeric data, textual descriptions, GIS maps, photos, video, satellite images, physical and mathematical models, and the knowledge, probably in the mind of expert scientists and modellers, of how they work.

Decision *(output)*	**Support** *(tools)*	**System** *(input)*
interpretation	database	data
integration	rule-base	text
presentation	graphics	maps
classification	statistics	photos
scenarios	expert system	video
cost benefit	neural network	image
risk analysis	uncertainty analysis	models
recommendations	optimization	knowledge

Figure 1. A new decision support system framework.

To convert the input into the desired output for decision making poses quite a challenge and requires a new, three-stage approach. First, the system must have the ability of storing, accessing, screening and combining the input. This stage may involve the storage and retrieval of time series data in a database, the text in some frame structure, the photos and images in some graphics component, the maps and polygonal information in a GIS subsystem, the models in a linkable data exchange and programming environment, and then the knowledge in a rule-base component. Second, the system needs knowledge-based tools such as expert systems to work on these data and information to bring forth the output results for decision making. These tools are the logical core of the whole system, without which there will be no intelligent interfaces and the system may be reduced into an ordinary environmental information system, like a GIS or a database. It is also this logical or inference engine component that enables the linkage of the quantitative and qualitative information. Because of this feature, we call this new approach the knowledge-based system approach. Third, when the logical procedures are established to use the different forms of input with these support tools, the procedures may have to be repeated iteratively many times for various scenario testing and cost analysis etc., using statistical, optimization or uncertainty evaluation techniques. The outcome must be prepared in simple tabular, graphical or textual forms for easy understanding. The implementation of the approach should provide the mechanism for the individual customarization of user-friendly interfaces as required by the decision makers. After all, the system should help the managers and their advisors, not hinder them.

3. IMPLEMENTATION: A PARADIGM SHIFT

With the advent of microcomputer technologies and the easy access to information on the Internet, there is certain expectation, at least in the mind of those decision makers and advisors who are aware of this paradigm shift in information management, to see the implementation of the environmental decision support system to be in line with the new techniques. They want the information and analysis results literally at their figure tip; they want to test scenarios and prefer to have the cost and benefit results almost instantaneously. The new framework shown in Figure 1 can meet these demands readily.

For example, over the past decade, we have been developing a generic, knowledge-based, environmental decision support system, acronymed RAISON for regional analysis by intelligent system on microcomputers, currently implemented on the Windows 95 platform (Lam et al., 1995). As a generic system, RAISON does not provide data, models, rules and knowledge. For each application, these must be entered from external sources first. It offers system input (Figure 1) facilities that accept various types of data, text, maps and images from external sources into its own internal, fully linked database, map subsystem and graphic components. Often, there are two phases in the construction of the decision support system: the technical user interface (TUI) and the public user interface (PUI). The TUI's are used by technical users for connecting databases, rule-bases, models and other information. The technical user can also use the support tools (Figure 1) including the logical components of RAISON to control and direct the decision processes. When completed, the TUI's will form the basis of the PUI's. The PUI's are for managers and stakeholders and may even be used in public consultation meetings (Young et al., 1997). They therefore need to be composed of simple and easy operations using pull-down menus, buttons and hypertext, especially customarized for the problem using the support tools (Figure 1). Because of its generic and flexible implementation, over twenty environmental applications have been constructed using RAISON (Lam et al, 1995). Here, to illustrate the main features of the new approach, we use the following five decision support applications based on the RAISON system: interpretation, advice, assessment, scenario testing and cost benefit analysis.

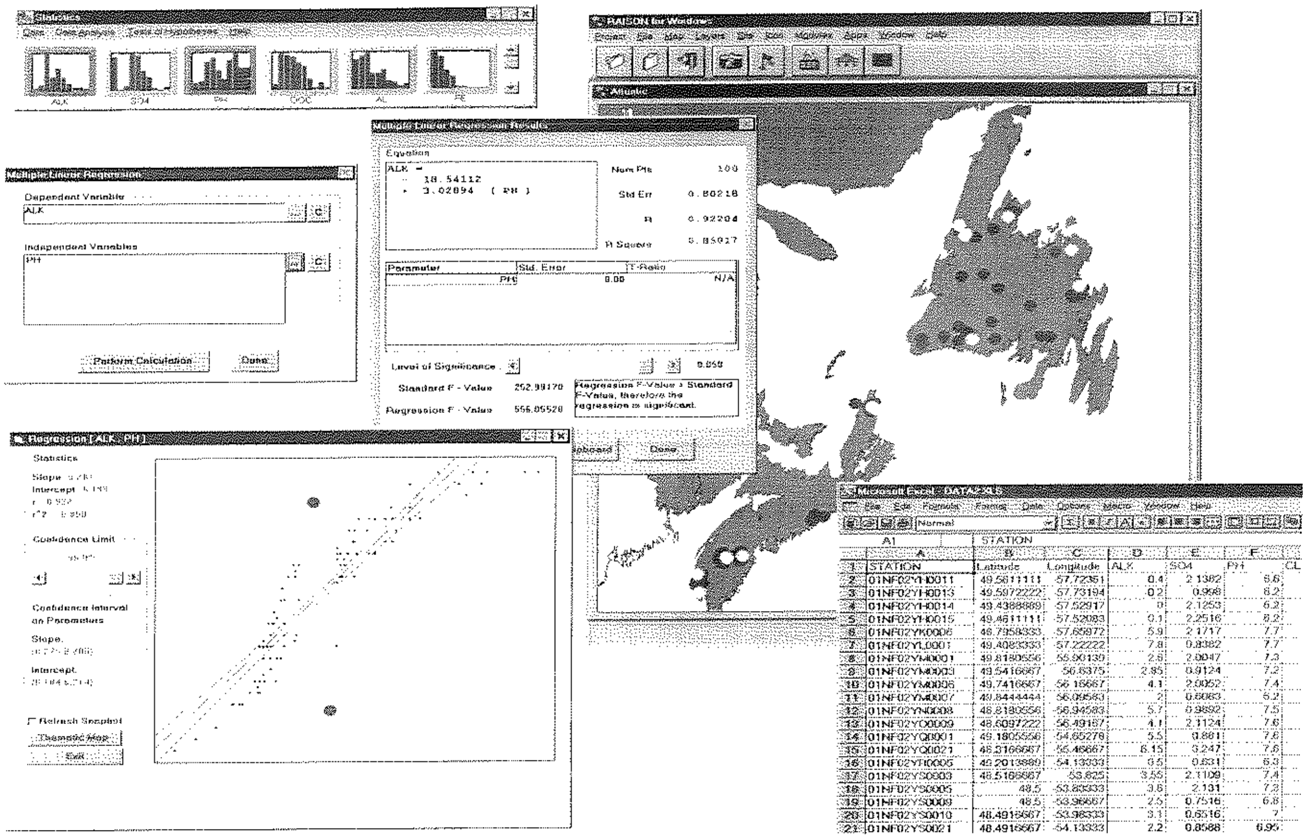

Figure 2. Clockwise from lower right: the EXCEL spreadsheet used to input data to RAISON; graph showing the regression analysis results with confidence bounds; the interface for multiple linear regression analysis; mini-graphs showing the statistical distribution of the parameters (ALK, SO4, etc.); the numerical results of the regression analysis; a map showing the locations for data within (white dots) and outside (black dots) the confident limit of the regression relationship between ALK and pH.

4. APPLICATIONS

4.1 Data Interpretation

Simple semi-empirical relationships between water quality and ecological variables usually play an initial and important role in supporting environmental decisions. The interpretation of data is often based on regression analysis and constitutes the beginning step towards more complex analysis. As an example, existing data from an EXCEL spreadsheet (Figure 2) were entered into the RAISON system via the worksheet component. A map (Atlantic Canada, Figure 2) showing the locations of sampling sites was also imported. Since the spreadsheet contain the station names and the latitudes and longitudes of the sites, the numeric data were automatically connected to the map. A quick review, for example, from mini-graphs (Figure 2) showing the statistical distribution, may help select which pair of variables to be the candidates for the further analysis. For instance, in Figure 2, the parameters, alkalinity (ALK) and pH, were chosen for regression tests, because the two distributions were fairly similar. The regression results, both graphical and numerically (Figure 2), confirmed this hypothesis, with the correlation coefficient value of $r^2 = 0.85$. The statistical results were displayed back on the map, e.g. by highlighting those sites with data that were within the 95% confidence limit. Subsequently this relationship between ALK and pH was used to characterize geochemical processes for this region required for acid rain impact analysis (Lam et al. 1989).

Figure 2 shows some fundamental features in a simple TUI: bringing data (spreadsheet, map) from external sources, searching quickly for candidates for further analysis, performing the analysis and displaying the results as simple tables, graphs or maps. While this may not be a PUI that supports important decision making, it does help scientists search relationships that could help decision making. The above procedures also form part of the first stage of constructing decision support system as discussed in Section 2.2. Note that these procedures can be also implemented outside of RAISON by using spreadsheets that can be connected to some GIS and statistical software, requiring the linking protocols to be written by the user. In the RAISON system, however, the linking protocols are automatically set up, once the data are entered. For development into a PUI, descriptive knowledge and intelligent interfaces are also required, as discussed next.

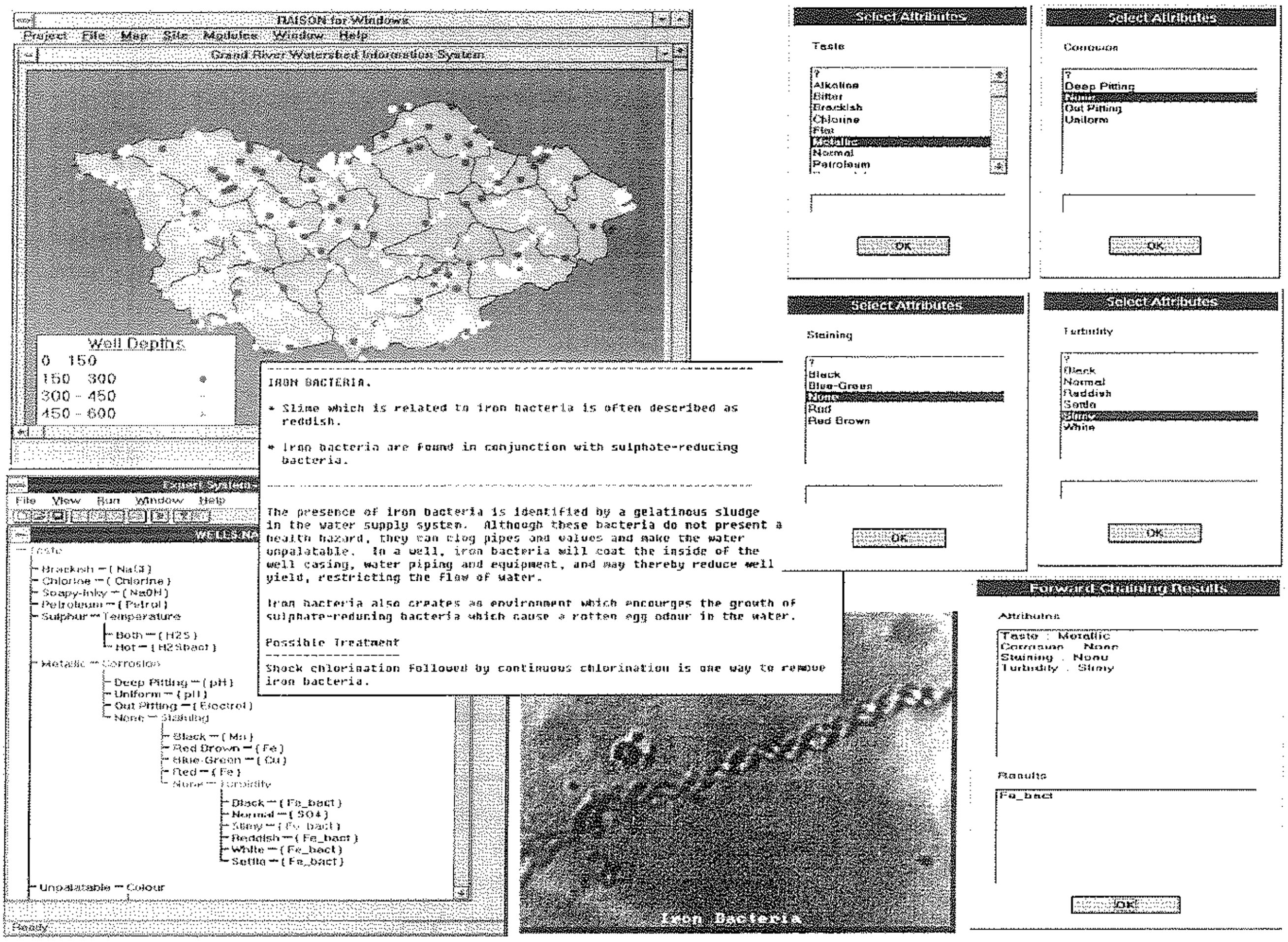

Figure 3. From bottom left: the decision tree structure representing the expert system rule-base for the drinking water well advisor; a map showing the wells; the query sequences on taste, corrosion, staining and turbidity; forward chaining results indicating the conclusion to be an iron bacteria problem; a photo of the bacteria; information on the bacteria and recommendation for possible treatment.

4.2 Expert System Advisor

In the knowledge-based approach, the logical component in a decision support system consists of some intelligent programs such as an expert system. An expert system is a computer program that stores descriptive knowledge, usually acquired from experts, as a rule-base (or frames, or other knowledge structures) and uses the knowledge via an inference engine to infer conclusions from hypothesis (forward chaining) or to evaluate hypothesis for given conclusions (backward chaining). In the RAISON system, the expert system can be used by itself and can be linked to databases, maps, models and other rule-bases. An example is the drinking water well advisor. In this application, the knowledge on the diagnostics of the colour, odour, taste and other characteristics of the drinking water obtained from groundwater wells were input as a rule-base. The rules were acquired by interviewing the experts and entered into the rule-base as a table in spreadsheet format, with each row constituting one 'if-then' rule. Once these rules were entered, they were converted into a tree structure (Figure 3). To use the expert system, a series of queries were asked and the answers were selected from a pull-down menu, until a conclusion was reached. Figure 3 shows a typical query session which led to the conclusion that the drinking water well contained iron bacteria, after four questions. In the RAISON system, each of the conclusions could be linked to a hypertext describing the diagnostics and the possible treatment so that the corresponding document, including a photo (Figure 3) or a video, can be displayed. The expert system can be activated by clicking at any of the well sites on a map individually, or can be used to process a given set of data (e.g. surveys) in a batch mode. The application can be used either as a TUI or a PUI and appears to have the intelligence of an expert, because it can diagnoses symptoms, reaches a conclusion and provides some good advice for the user. In advanced applications, the expert system can be used to select the most appropriate model from a collection of models for given conditions, or the rules can be modified to allow for marginal overlapping and conflicts using fuzzy logic. These are all essential for the second stage of system development (Section 2.2).

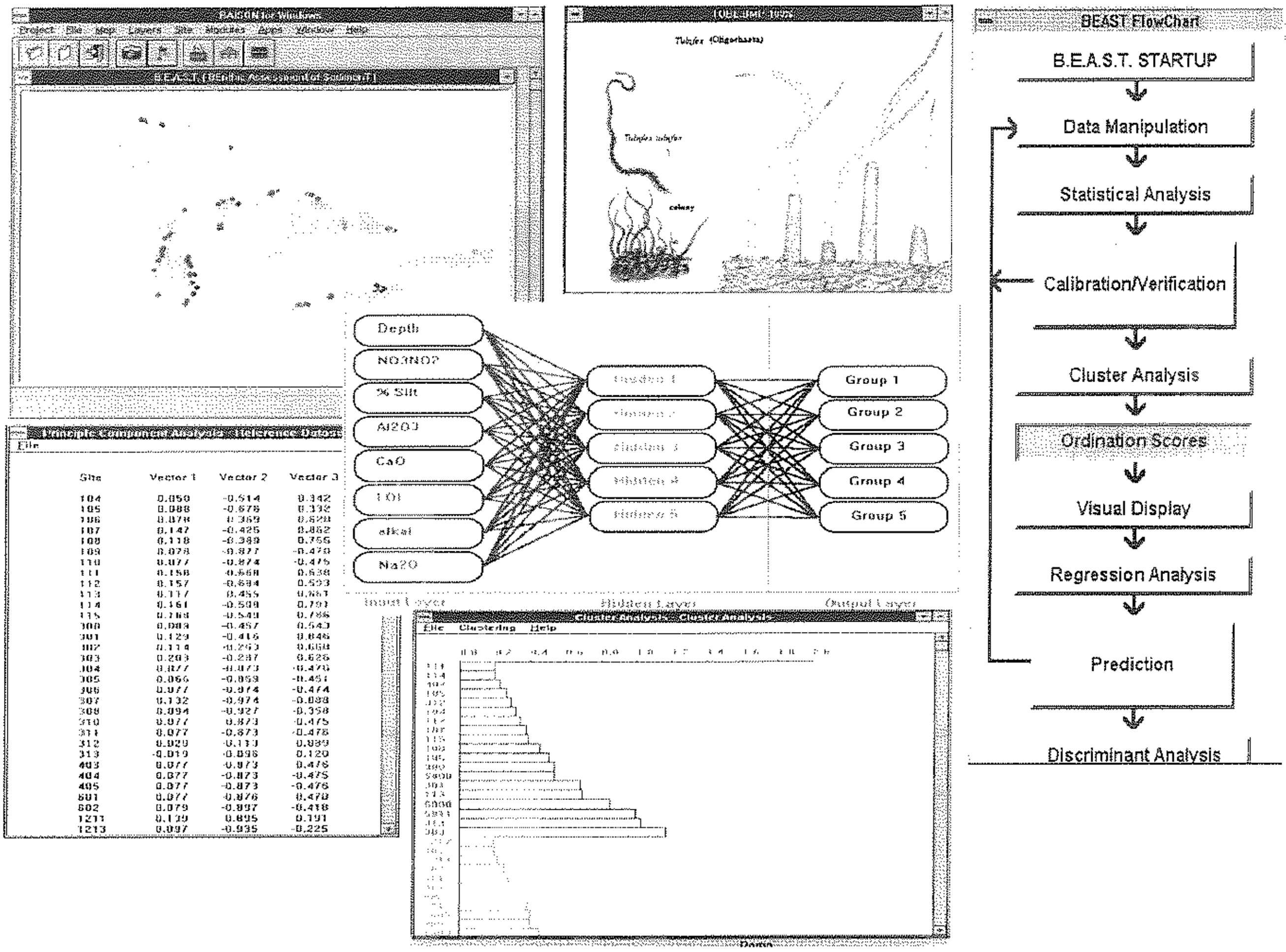

Figure 4. Clockwise from top left: the North America Great Lakes nearshore sediment data and sites; a photo of some benthic organisms; a flow chart to guide the use of the decision analysis; graphics showing classification grouping by cluster analysis; a table of principal component analysis results by sites; neural network for site classification using environmental data.

4.3 Assessment and Classification

Environmental assessment concerns a broad spectrum of impact analysis, policy guidelines and resources at risk. The use of ecological indicators as a means of establishing assessment guidelines has become an important aspect in decision making. An example is the classification and identification of sites for clean-up or remedial actions according to the environmental conditions in support of the well-being of benthic organisms in sediments for lakes and rivers. For example, Reynoldson et al. (1995) proposed to use multivariate techniques involving data on the biological structure of benthic invertebrate community, functional (survival, growth and reproduction) responses from bioassay tests, and environmental water and sediment quality data. The methodologies involve cycles of statistical tests such as regression analysis, cluster analysis, principal component and discriminant analysis, etc. The difficulties with such an approach included the choice of statistical methods, the connection of the methods to environmental databases and the presentation of the results by sites or zones on a map for subsequent clean-up actions. To circumvent these difficulties, a TUI was constructed in the RAISON system that provided direct linkages among statistical methods, database, maps and graphical support. A special interface, with a flowchart of the methodologies, guided the user to follow various steps for calibration, analysis and prediction (Figure 4). In addition, RAISON provided the option of using neural network tools for classifying sites by different levels of contamination and clean-up actions. The neural network (Lam and Swayne, 1996) is another powerful tools required by the knowledge-based system approach (Figure 1) for knowledge extraction and manipulation in a decision support system. The neural network is based on the neuron structure for acquiring knowledge, e.g. learning patterns and trends from a training set of input and output data, using parallel and feedback mechanisms with hidden layers (Figure 4). When the neural network is trained, it can be used for prediction with new input data. In this example, the neural network gave very similar results as the statistical methods. A PUI was subsequently developed based on the TUI information.

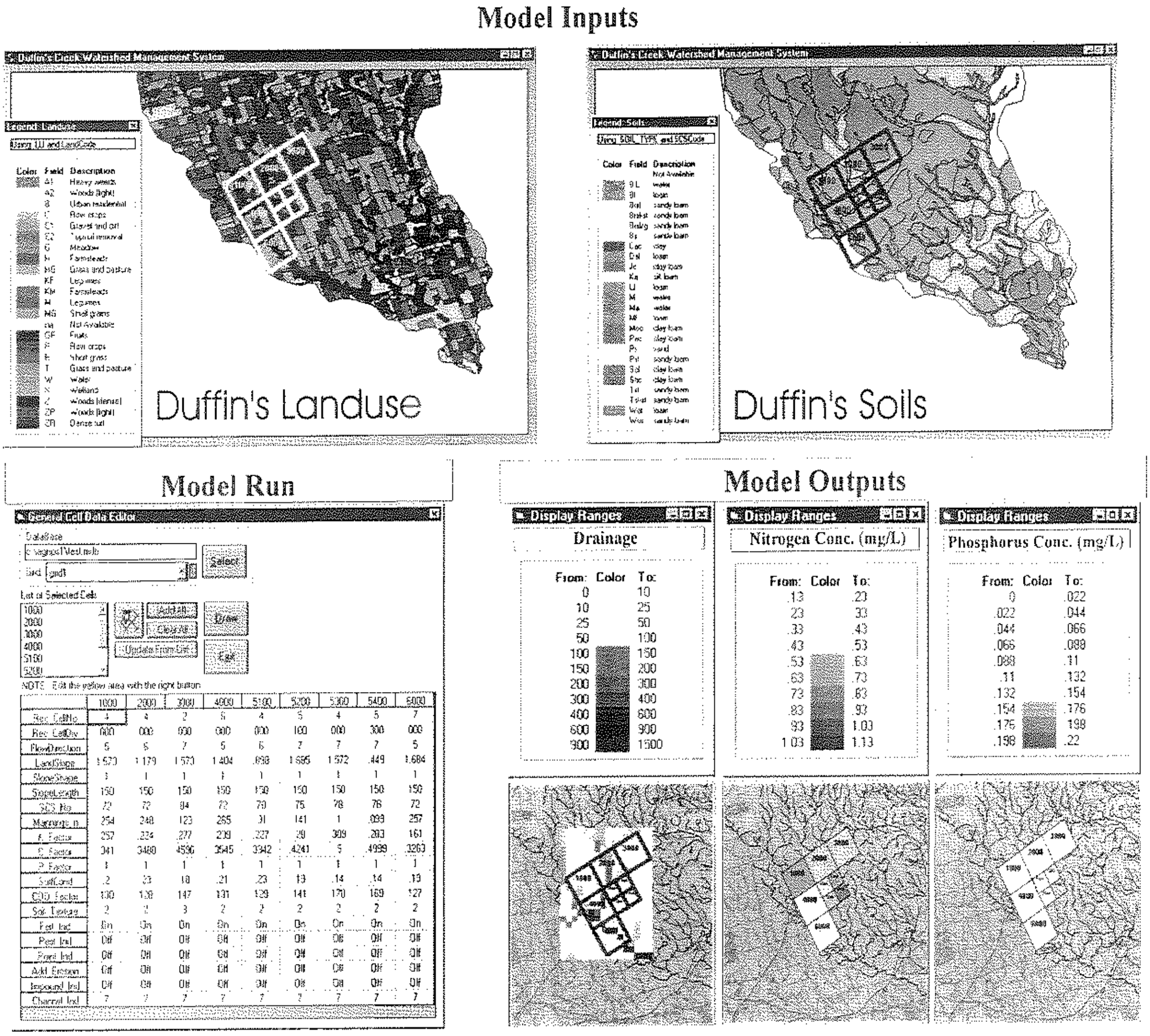

Figure 5. Clockwise from top left: Polygonal information from a landuse map is used for defining model coefficients over a grid system for a nonpoint pollution model; same as previous window but for soil types; model output on flow drainage, nitrogen and phosphorus concentrations highlighted on grid cells; model run using coefficients as defined by the maps and other input.

4.4 Scenario Testing

Many decision processes often require some forms of scenario testing or gaming, especially on 'what if' questions. Mathematical models, when verified and validated, are commonly used to generate outcomes due to altered input or boundary conditions for decision support purposes. An example is the use of the Agricultural Non-point Source Model (AGNPS) for predicting pollutant concentrations for watershed management (Young et al., 1989). To embed a mathematical model in a decision support system, however, is a nontrivial exercise. Either the model has to be rewritten in a programming language acceptable in the decision support system, or to keep the model coding as is but linking input and output files through such techniques as dynamic data exchange, dynamic linkage library and object linking and embedding. For example, in Lyon et al. (1997), the AGNPS model coding was kept intact, with linkages to RAISON database, map layer, object and graphics components. The result was a user-friendly TUI interface (Figure 5) that could directly convert information from landuse, soil type and topography maps into values of model coefficients, and, after the model was run, the results were presented on the map effectively for considerations by watershed planners. Because of these direct linkages, the input from the maps could be altered easily, for example, by changing part of croplands into paved surfaces (i.e.,what if there is urban development), or all urban and agricultural lands into forests and wetland (i.e., what if we go back to pristine conditions). The model coefficients were then updated automatically and new model results displayed immediately. In the past, map alterations and model re-definition may take days, but now it requires only minutes to change the scenarios and to obtain new answers.

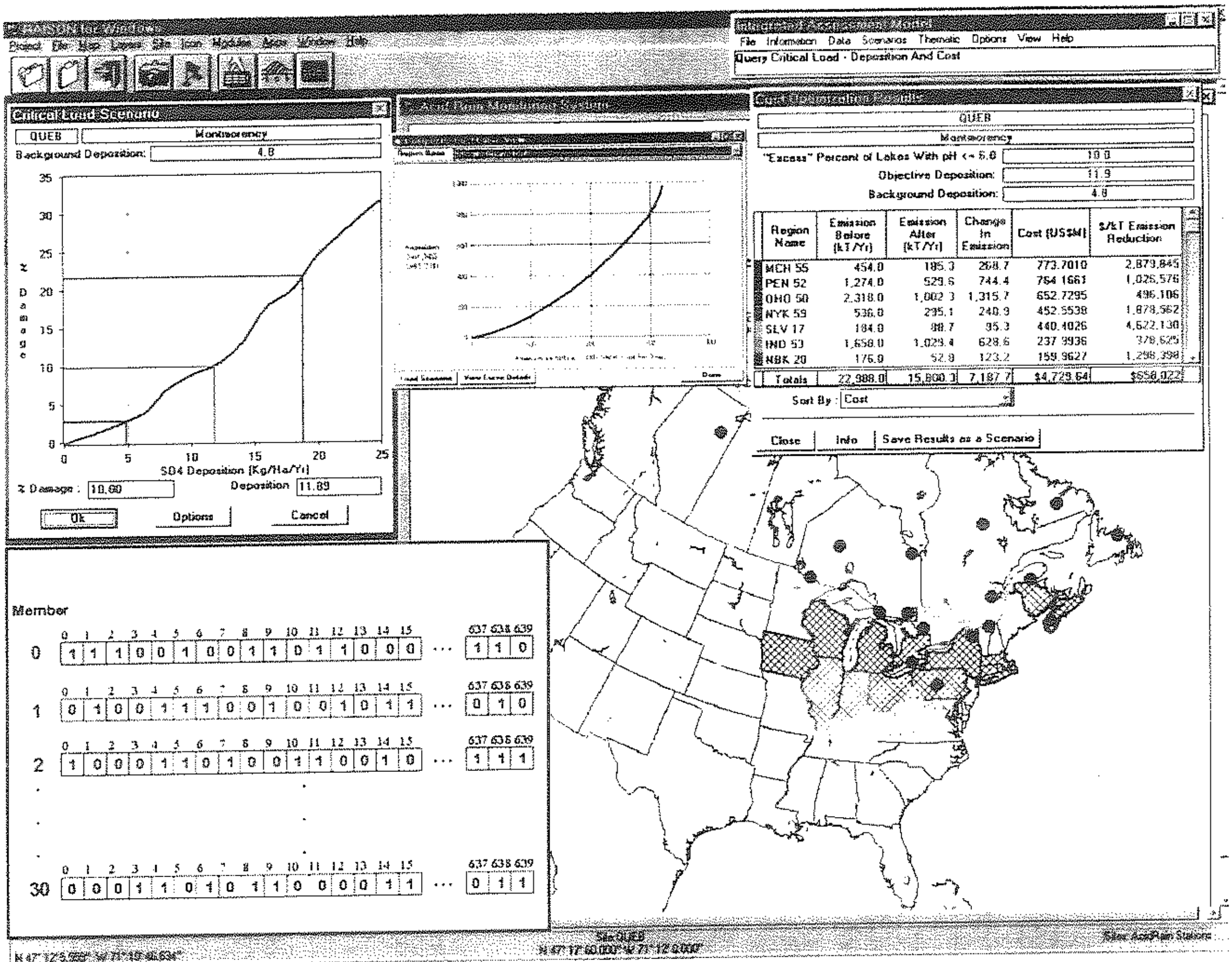

Figure 6. Clockwise from top left: critical load scenario curve for Montmorency, Quebec, Canada; cost function curve for emission reduction in Mucigen, U.S.A.; table of cost optimization results; map showing source regions affected in the optimization; and iterative mechanisms used in the genetic algorithm (normally not shown on PUT).

4.5 Cost Benefit Analysis

Almost all decision making processes eventually involve costs. Environmental problems are no exception. To combine scientific and economic data and models, however, poses a new challenge, as the two domains are often not intimately related, and therefore requires support linkages described above to construct a decision support system that enmeshes the knowledge from both domains. An example is the cost benefit analysis for reducing acidifying emissions in North America (Lam et al., 1997). Figure 6 shows the critical load curve for wet sulphate deposition at the Montmorency receptor site near Quebec City, Canada. This curve represents a summary of the knowledge of the scientific domain. It was constructed by first running several geochemistry models and then the model results were screened by an expert system (Lam et al, 1989). When this expert system model was run repeatedly with incremental sulphate deposition values, different predicted levels of pH and hence lake damage were obtained. This critical load curve was used in a PUI in RAISON by plotting the damage level versus the deposition. For example, from Figure 6, by moving the cursor to the 10% lake damage level, the estimated sulphate deposition was shown to be about 11.9 kg/Ha/yr, a drop from the current value of about 22 kg/Ha/yr. The question is: what should the minimal cost be and where should the emission reductions be made? To answer this, we need the information on reduction costs for all emission sources (e.g., Figure 6 shows the cost function for the source region of Michigan). To relate the lake damage curves to the cost function, one needs the source to receptor transfer matrix model (Lam et al. 1997), relating the emissions to the deposition. Once these models were connected, one then called for the optimization procedure. The algorithm used for the cost optimization was the genetic algorithm (Goldberg, 1989) which apparently had a better searching strategy than most nonlinear programming procedures by reserving acceptable portions (genes) of the iteration results (Figure 6). For the 10% lake damage level as chosen, the total optimal cost is about \$4.73 billion U.S. dollars (1995 value) per year for a total sulphur dioxide emission reduction of about 7260 kT/year. The individual reductions and their costs were also estimated as shown in the cost optimization results table (Figure 6), with the Michigan source region bearing the highest cost in this example (the results shown here were for illustrative purposes only). This is also an example of the third stage of developing a decision support system(Section 2.2). When all three stages are completed, the manger has a PUI with all the necessary scientific and economic information to make decisions.

5. DISCUSSIONS AND CONCLUSIONS

While there are many more case studies, we have sufficiently illustrated several key features in an environmental decision support system with the five examples. Theses examples are mainly of the technical user interface (TUI) type, because all of them were developed from scientific findings first and then evolved into a nontechnical public user interface (PUI) for managers and stakeholders. However, if a system is initially developed with a strong management input, it may have the advantage of being more effective for decision making as it is what the manager asks for, provided that the science that is added later is fundamentally sound. On the other hand, a system developed without good science is obviously not too useful. Therefore, the incorporation of appropriate scientific knowledge for a given problem is a very crucial step. A simple semi-empirical relationship is acceptable as long as it provides the correct answer. A decision support system that helps scientists search and define such relationships, whether using regression analysis or neural network, is as good as one that uses complex models or advanced statistical techniques. For simple decision support systems, the control of the logical flow can be easily handled by a menu or an effective interface. For more complex situations, e.g. one that links many models for cost benefit analysis, it is better controlled by an expert system with support of databases, maps and objects. Implementation of such a system can be made as one specific application program or as a generic system. We have opted for the generic approach, because it can be adapted easily from one application to another. We have built the generic system with many applications. We have found in all these applications that there is always some new practical features to be incorporated into the generic tool-kit for use in the next application.

In conclusion, we have presented the knowledge-based system approach to the development of decision support system for environmental and water resources applications. It was found that an intelligent core is essential for combining knowledge from different domains. In complex problems, reliable qualitative knowledge helps fill the gap left by quantitative knowledge. Also, with the increase in the ability of accessing data and application programs through the Internet, it is expected that future decision support systems will be made available on the information highway. Further research, however, is still required for describing the uncertainty of these systems and for post-auditing the outcomes of scenarios. Only when the uncertainty and predictability are confirmed would managers and scientists have the confidence needed to promote and encourage the use of these systems.

Acknowledgements. The author is indebted to the RAISON development team for suggestions and advice and to the many collaborators and research partners for data and knowledge in writing this paper. He thanks D. Swayne, A. Storey, K. Brown, W.G. Booty, I. Wong, P. Fong, G. Bowen, L. Lyon, M. Moran, T. Reynoldson, D. Jeffries, A. Mariam and others in contributing to the examples used.

References

Goldberg, D.E.Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA., U.S.A., 1997.

Lam, D.C.L., Swayne, D.A., Storey, J. and Fraser, A.S. Watershed acidification using the knowledge-based approach, *Ecol. Modelling* 47, 131-152, 1989.

Lam, D.C.L., Booty, W.G., Wong, I., Kay, D. and Kerby,J. RAISON decision support system: examples of applications, *NWRI Contribution Rep.* No. 95-151, Burlington, Ontario, Canada, 23p., 1995

Lam, D.C.L. and Swayne, D.A. A hybrid expert system and neural network approach to environmental modelling: GIS applications in the RAISON system. *IAHS Publ. No. 235*, 685-693, 1996.

Lam, D.C.L., Puckett, K.J., Wong, Moran, M.D., Fenech, G., Jeffries, J.S., Olson, M.P., Whelpdale, D.M., McNicol, D., Mariam, Y. K.G. and Minns, K. An Integrated Acid Rain Assessment Model for Canada: from source emission to ecological impact. Wat.Quality Res. . Can. (submitted 1997).

Leon, F., Lam,D.C.L., Swayne, D.A., Farquhar, G.J. and Soulis, E.D. Integration of nonpoint source pollution model with a decision support system. Proc. 2nd Int. Symp. Environmental Software Systems, Whistler, B.C., Canada, in press, 1997.

Reynoldson, T.B., Bailey, R.C., Day, K.E., and Norris, R.H. Biological guideline for freshwater sediment absed on Benthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state, Australian Journal of Ecology 20, 198-219, 1995.

Young, R.A., Onstad, C.A., Bosch,D.D., and Anderson, W.P. AGNPS: A non-point source pollution model for evaluating agricultural watersheds. *J. Soil and Water Conservation* 44, 168-173, 1989.

Young, W.J., Lam, D.C.L., Ressel, V. and Wong, I Development of an environmental flows decision support system. Proc. 2nd Int. Symp. Environmental Software Systems, Whistler, B.C., Canada, in press, 1997.