Application of Knowledge-based Tools in Environmental Decision Support Systems

David C.L. Lam and David Swayne and Yohannes Mariam and Isaac Wong and Philip Fong

1997

[1]Application of Knowledge-based Tools in Environmental Decision Support Systems

David C.L. Lam[1], David Swayne[2], Y. K.G.Mariam[3], Isaac Wong[1] and Philip Fong[1]
[1]Environment Canada, P.O. Box 5050, Burlington, Ontario, Canada L7R 4A6
[2]Dept. of Computing and Information Sciences, U. of Guelph, Guelph, Ont., Canada N1G 2W1
[3] Regulatory and Economic Assessment Br., Environment Canada, Ottawa, Canada K1A 0H3

ABSTRACT

Decision support system often requires the combined knowledge of multiple domains. A knowledge-based approach is proposed to include not only the process modelling knowledge but also the descriptive knowledge in the integration. Descriptive knowledge such as survey statistics and expert opinions forms the core of a study on the uncertainty of the combined knowledge. It was found that the use of expert systems, neural network and belief causal network assist greatly in the implementation of these concepts. Examples are drawn from the combination of scientific and economic knowledge to solve some acid rain problems.
Keywords: decision support system, knowledge-based system, expert system, causal network

INTRODUCTION

Environmental problems often involve many branches of sciences including physics, chemistry and biology for various media such as air, soil and water. Their solutions require the integration of knowledge from these scientific disciplines, sometimes in the form of simple semi-empirical regression relationships and at other times as complex mathematical models. Environmental problems also involve cost benefit analysis. Decision makers look for solutions that could protect the environment at a minimum cost to industry and society.

To find these solutions, we need highly sophisticated software tools not only for analyzing numeric data, but also for utilizing the non-numeric or descriptive knowledge in both the scientific and economic domains. In the past, while the descriptive knowledge was deemed essential, they were often overlooked until the advent of knowledge-based systems (Lam et al., 1996). These decision support tools can access and integrate numeric data, map information, mathematical models and knowledge-bases, and can optimize the economic costs under given constraints of environmental guidelines or ecological standards. Recently, we have embedded the development of knowledge-based systems with the requirements of environmental decision support systems (Lam et al., 1996). The models and expert system rule-bases are interconnected, through a built-in database and an internal geographical information system (GIS), with the capability to select models and data using artificial intelligent tools such as expert systems and neural network. In this paper, we further review such technological advances in the context of environmental and economic analysis and the associated uncertainty.

KNOWLEDGE FOR DECISION SUPPORT: A NEW PARADIGM

Traditionally, the treatment of scientific knowledge has mainly been focussed on the numerical aspects of

---

Published in Proc. 8th Int. Conf., InterSymp '96, Systems Research, Informatics and Cybernetics, IIAS, Baden-Baden, Germany, (1996).

data. By constructing temporal trends and spatial patterns from data, one usually gains the insight and explanation through such tools as statistics, modelling and simulation. When these models are calibrated, verified and validated, they are then used for making environmental predictions. To make use of the process knowledge in these models, however, such as the selection of appropriate models for a given set of data or the assignment of appropriate values for model coefficients for given initial or boundary conditions, we require additional knowledge. This extra knowledge is usually descriptive and may be based on information gained through

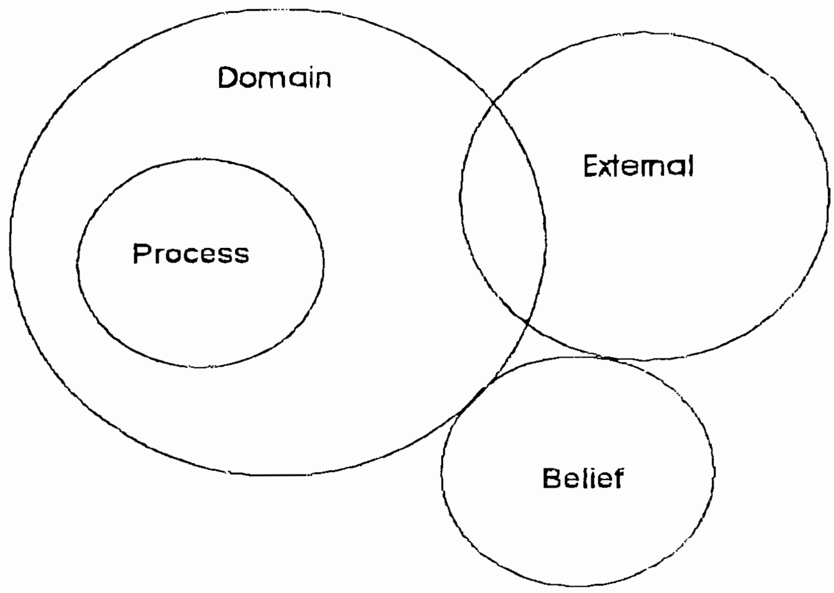

Figure 1 Knowledge framework in knowledge-based systems.

experience or familiarity with the domain knowledge (Fig. 1). Thus, the domain knowledge includes the process knowledge and the knowledge to utilize it.

The requirement for descriptive or qualitative knowledge is also essential when one needs to connect the knowledge of an external domain (e.g. economics) to the scientific domain to solve a problem. The external domain (Fig. 1) may have numeric process knowledge such as economic cost functions or economic forecasting models. In order to connect process models of one domain to process models of another, the descriptive knowledge of the two domains need to be combined. There may be overlapping areas in these two domains, but the overlapped areas are likely to be few (Fig. 1) because, for disciplines that are relatively far apart, experts in both field would seldom work together. Therefore the knowledge of the overlapped areas (Fig. 1) is small. If the common knowledge between these two domain is limited, the uncertainty in the combined models and descriptive knowledge will be even more so. On the other hand, this limitation should not prevent experts from postulating or guessing the uncertainty, based not so much on available data but on expert beliefs or opinions. While it may sound unusual for scientific disciplines, beliefs or survey opinions are used frequently in socio-economics. Recently, belief causal network analysis (Jensen et al., 1990) has been applied for environmental problems and is a new powerful tool to investigate knowledge domains where available knowledge is weak but the expert opinions can be counted on as a good, first guess. This belief knowledge is often not part of, but is peripheral to, the combined knowledge (Fig. 1), unless the belief is proven to be valid.

EXAMPLE I: COST OPTIMIZATION AND THE GENETIC ALGORITHM

In Lam et al. (1996), we presented the integrated assessment modelling approach to combining scientific knowledge for processes in air, water, soil and ecology to predict the environmental impact due to acidifying emissions. These tools

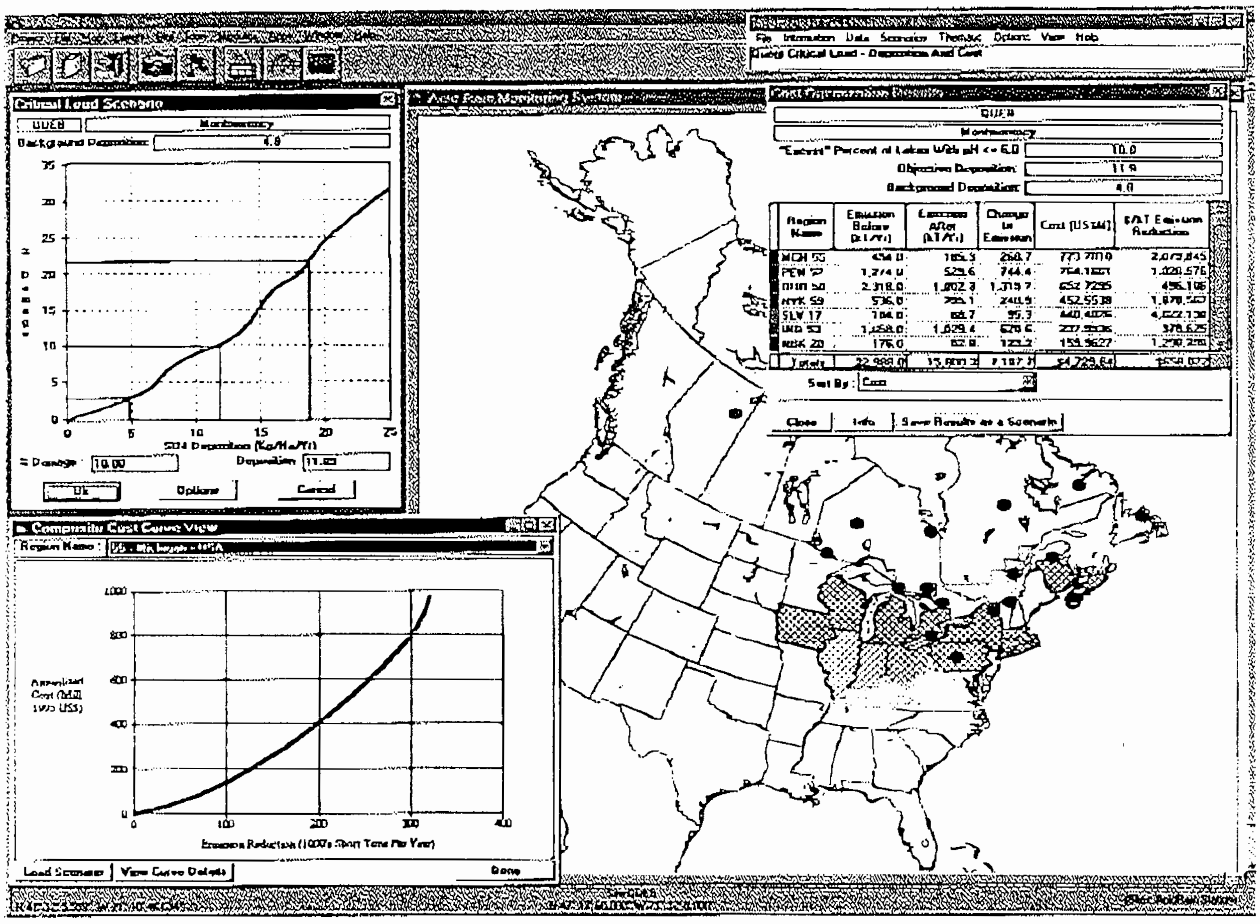

Figure 2 Clockwise from top left: critical load scenario curve, cost optimization results, map showing source regions affected in optimization and cost function curve.

were implemented as part of the RAISON (Regional Analysis by Intelligent Systems ON microcomputers) for Windows system (Lam et al, 1996) designed for decision support applications. In this paper, we continue to use the RAISON system to further illustrate the knowledge-based paradigm (Fig. 1) by linking these scientific results to economic analysis. Figure 2 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 to find the most appropriate predicted values of pH for a given set of soil sensitivity and water quality data (Lam et al, 1996). When this expert system model was run repeatedly with incremental sulphate deposition values, different predicted levels of pH and hence lake damage were obtained. By graphing the damage versus the deposition, the critical load curve was constructed. For example, from Figure 2, at the 10% lake damage level, the estimated sulphate deposition was 11.89 kg/Hacres/year.

To demonstrate the cost optimization procedure, suppose that we accept this lake damage level and critical load as the target objective. The next question is: what is the least-cost emission reduction strategy that would enable the attainment of predetermined ecological targets? To do so, we need first the information on reduction costs. These costs were estimated for all forty emission source regions in Northern America (Fig. 2). An example of the cost function estimated for the source region of Michigan, was shown in Figure 2. Note that the cost function, as in most cases, was a nonlinear function which increases rapidly when further reduction may require more costly technologies. To relate the lake damage curves obtained from the environmental domain to the cost function obtained from the economic (i.e., external, Fig. 1) domain, one needs to connect the process models. In the integrated modelling approach, these vital pieces of information were stored in the knowledge base and can be retrieved through the graphic interfaces as shown. Once they are connected, one can call for the optimization procedure via the option button provided at the bottom of the critical load scenario interface (Fig. 2). This would then invoke another interface (not shown) that deals with optimization. It will ask for an upper bound for the percentage reduction at each source region. Sometimes, the choice of the target objective, i.e. lake damage level, may be so low that optimization is not possible. The interface will issue a descriptive warning about it and ask for another trial, possibly by relaxing the target objective, increasing the upper bound for reduction or both. It is this communication about the descriptive knowledge that makes this knowledge-based approach more informative and effective than other approaches. Since the cost function is nonlinear, ordinary linear programming will not be applicable and nonlinear programming techniques are required. The algorithm used for the cost optimization was the genetic algorithm (Goldberg, 1989) which apparently has a better searching strategy than most nonlinear programming procedures. To bridge the target objective specified as lake damage level and least-cost emission abatement option, however, a source-receptor matrix (Lam et al., 1997) was used. From the specified lake damage level, the corresponding sulphate deposition was determined (Fig. 2). For this sulphate deposition, the source-receptor matrix was used in the genetic algorithm to generate the required emission reduction levels. Normally the source-receptor matrix was used to predict the deposition at a receptor site for a given array of emission sources. Since the matrix was linear, it could be used to predict the total emission reduction required for a decrease in deposition at a given site. The genetic algorithm was used to assure that this total reduction amount was correct, while searching for the optimal individual amount to be reduced at each of the source regions. Figure 2 shows an example of the outcome of the cost optimization. For a 10% lake damage level, which represented a substantial improvement from the current level of about 22% (Fig. 2), the total optimal cost is about $4.73 billion U.S. dollars (1995 dollar 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 (Fig. 2), with the Michigan source region (see cost curve, Fig. 2) bearing the highest cost (US $773.7 millions) in this example (the results shown here were for illustrative purposes only).

EXAMPLE II: MODEL UNCERTAINTY AND THE CAUSAL BELIEF NETWORK

Within the integrated assessment model, there are uncertainties associated with model input and coefficients, as well as uncertainties in the observed data. While techniques such as expert system (Lam et al., 1996) may help minimize the uncertainties in the model choice, there exists an even more complex problem when component models from different disciplines are combined. The uncertainty propagation through the linkage can at best be treated as a probability exercise, given that there are many unknown processes governing the combination of uncertainties. Conventional uncertainty analysis methods such as Monte Carlo process can be computationally prohibitive. Therefore, we propose to use the causal network approach (Jensen et al., 1990), since it can be based model variables or inputs and a set of directed links between them, a situation well suited for the various linked models as discussed in Example I.

For example, consider the uncertainty in the source-receptor matrix that linked the lake damage level to the cost function in Example I. This link was affected by three types of uncertainties: input uncertainty (I), inter-variable process uncertainty (V) and uncertainties associated with spatial interpolation (K). The uncertainty for deposition prediction at any receptor has the combined uncertainty (M) of these three factors, as well as the uncertainty of background and dry deposition (S). Figure 3 shows the causal network for the receptor site at Kejimkujik, Nova Scotia, as an example, with five main source regions considered in Canada, five in the U.S. and one for other areas. Note that for each major source region, there were five variable nodes (I, V, K, M, S) linking eventually to the receptor site. The probability for high uncertainty was, from statistics and expert opinion, in the range between 10% to 30% for the variables I, V and K. We combined probabilities by using the relation P(A|B).P(B) = P(A,B), where P(A,B) is the joint probability of events A and B. Thus, I, V and K were combined into an aggregate probability (M) of about 26 to 30% for high uncertainty from these source regions to the receptor site at Kejimkujik. When the probability for high background uncertainty S (15% to 25%) was added, the probability for high uncertainty was combined to about 30% for the five main Canadian source regions, 31% for the five U.S. source regions and 29% for the other source regions. These were in turn combined finally into a 37% probability for high uncertainty in the predicted total sulphate deposition for Kejimkujik.

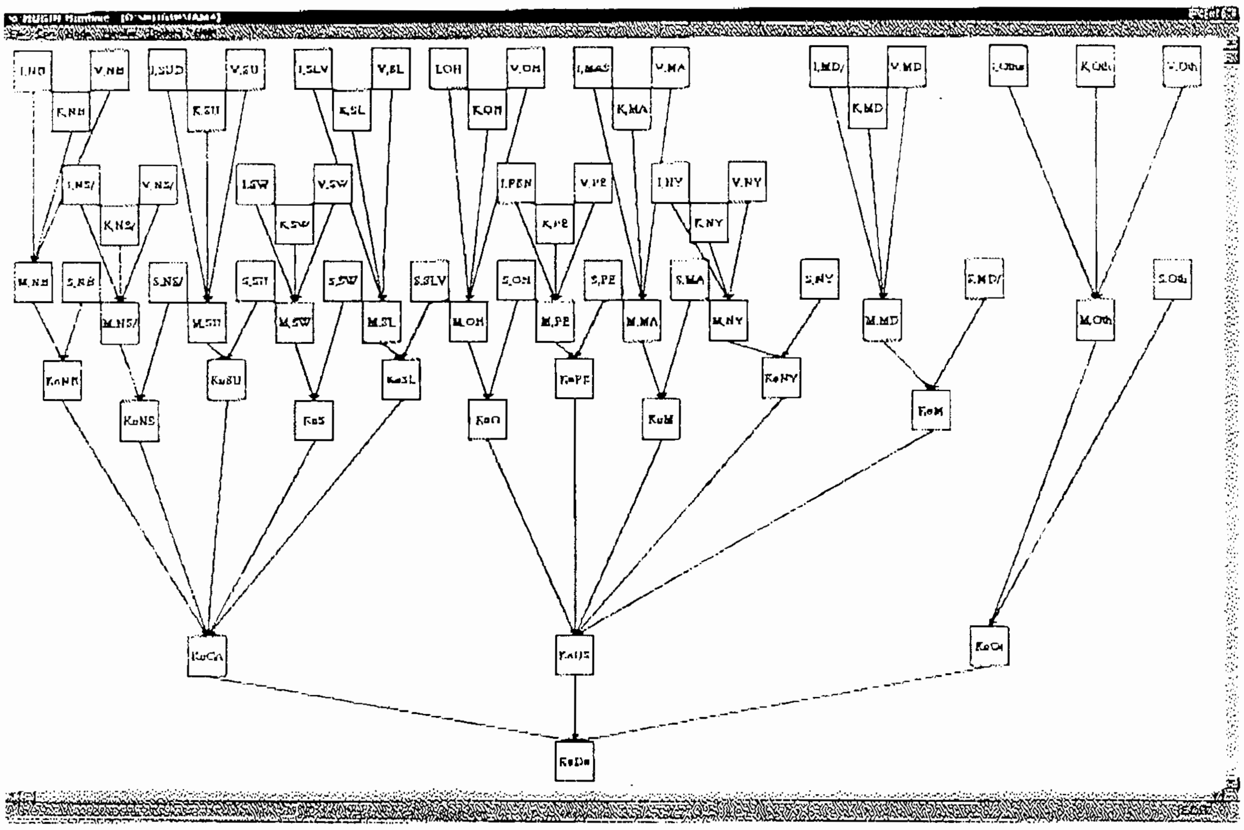

Figure 3 Belief causal network for the source-receptor matrix.

CONCLUSIONS

The knowledge domain approach to decision support goes beyond traditional process modelling. It requires the descriptive knowledge in the linkage of models from different knowledge domains and provides advice to avoid inconsistency and incompatibility. Preliminary results connecting environmental and economic domains are promising. Further work is required to explore this new concept, particularly in the area of model uncertainty.

Acknowledgement. A. Storey and K. Brown helped in the GUI design. Thanks are also due to the team support from D. S. Jeffries, M. Barre, K. Hedley, M. Moran, G. Fenech, A. McMillan and others.

REFERENCES

Goldberg, D.E. 1989. Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA., U.S.A.

Jensen, F.V., Chamberlain, B., Nordahl, T. and Jensen, F. 1990. Bayesian updating in causal probabilistic networks by local computation. Computational Statistics Quarterly, 4:269-282.

Lam, D.C.L., Swayne, D.A., Wong, I. and Fong, P. 1996. Advances in environmental information system: expert System and neural Network Applications. in Lasker, G.E. (Ed.): Advances in databases and expert systems, Vol. III, IIAS Publications, Baden-Baden, Germany, pp. 1-5.

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. 1997. An Integrated Acid Rain Assessment Model for Canada: from source emission to ecological impact. Submitted to Water Quality Res. J. Can.