International Journal of Environment and Pollution, Vol. 28, No. 3-4, 2006, pp. 297-309

Development of an integrated knowledge-based system on flow and water quality in Hong Kong coastal waters

K.W. Chau
Department of Civil & Structural Engineering, Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong
(email: cekwchau@polyu.edu.hk)

Abstract

This paper presents the coupling of the recent advancements in artificial intelligence (AI) technology with existing numerical models to constitute an integrated knowledge-based system (KBS) on flow and water quality. A hybrid application of the latest AI technologies, namely, KBS, artificial neural network, and, fuzzy inference system, in this specific problem domain is adopted here. This prototype system, serving both as a design aid as well as a training tool, is able to allow hydraulic engineers and environmental engineers to become acquainted with up-to-date flow and water quality simulation tools, and fill the existing gaps between researchers and practitioners in the application of recent technology in solving real prototype problems in Hong Kong. Moreover, the system can meet the demand for an integrated system that can quickly assist policy-makers in reaching decisions and also furnish convenient and open information service on water quality for the general public.

Keywords: integrated knowledge-based system; water quality; flow; Hong Kong waters; numerical model

Biographical notes

Dr. K.W. Chau is currently Associate Professor in Department of Civil and Structural Engineering of The Hong Kong Polytechnic University. He is very active in undertaking research works and the scope of his research interest is very broad, covering numerical flow modeling, water quality modeling, hydrological modeling, knowledge-based system development and knowledge management.

Introduction

It is essential to accurately and efficiently predict water levels, currents, and the transport and dispersal of pollutants in coastal waters. During the past decades, a variety of mathematical modelling tools employed as predictive tools in real prototype cases are developed. However, the emphasis on conventional computer-aided decision-making tools has been primarily placed on their algorithmic processes, in particular on the formulation of new models, improved solution techniques and effectiveness [1]. It should be noted that the selection of a suitable numerical technique and design parameters are influenced by many factors, such as water depth, water velocity, grid spacing, etc. The current technique for the numerical simulation of flow and water quality has become a highly specialized task, which can only be manipulated by experienced engineers. This has produced significant constraints on the use of models, thus generating a discrepancy between the developers and users of models.

The models are often not user-friendly enough. They lack the ability to transfer knowledge in the application and interpretation of the model, expert support for novice users, and effective developers-to-users communication. Many users of a model do not possess the requisite knowledge to glean their input data, build algorithmic models and evaluate the results of their model. The result may be the production of inferior designs and the under-utilization or even total failure of these models. Recently, as a result of the relatively low utilization of models in the industry, there has been an increasing demand for an integrated approach, which is also the current trend in the management of various systems.

Thus, the problem is to represent the information, knowledge and experience in such a format as to facilitate comprehension by a broad range of users from novices to experts [2]. It is worth investigating how to systemize the decision-making procedures of experts in simulating flow and water quality in coastal waters. This need is reinforced by recent advances in artificial intelligence (AI) and information technologies, which have rendered it possible to incorporate the heuristic knowledge into conventional algorithmic models. The information revolution of the last few decades has fundamentally altered the traditional planning, modelling and decision-making methodologies of water-related sciences and technologies. Information technology now plays an essential role in the sustainable development of water resources and the responsible management of the aquatic environment. In addition, the general availability of sophisticated personal computers with ever-expanding capabilities has given rise to increasing complexity in terms of computational ability in the storage, retrieval and manipulation of information flows.

This paper presents the coupling of the recent advancements in AI technology with existing numerical models to constitute an integrated knowledge-based system (KBS) on flow and water quality. An artificial neural network (ANN) for training of water quality parameters and a fuzzy rule base for the representation of heuristic knowledge are incorporated. This prototype system, serving both as a design aid as well as a training tool, is able to allow hydraulic engineers and environmental engineers to become acquainted with up-to-date flow and water quality simulation tools, and fill the existing gaps between researchers and practitioners in the application of recent technology in solving real prototype problems in Hong Kong.

Hydroinformatics

Hydroinformatics, which only received its proper name in 1991 [3], is a new and emerging technology and has since then become one of the most important branches of research and application in hydraulics and water resources. The new technologies with the widest applicability in the field include KBSs, ANNs, fuzzy systems, data mining, genetic algorithms and genetic programming [4]. Individual applications of these innovative techniques have recently been recorded in the literature. Chau and Chen [5] delineated an example of expert system on numerical modelling system in coastal processes. Zou et al. [6] applied an ANN embedding the Monte Carlo approach for modelling water quality under input information uncertainty. Chau [7] forecasted the river stage with particle swarm optimization ANN. Chen and Mynett [8] employed data mining techniques and heuristic knowledge in modelling the fuzzy logic of eutrophication in Taihu Lake. Cheng and Chau [9] devised a fuzzy iteration methodology for reservoir flood control operation. Mulligan and Brown [10] used genetic algorithms to calibrate water quality models. Chau [11] calibrated flow and water quality modeling by using genetic algorithm. Chen [12] evaluated the trophic state of reservoirs by applying genetic programming. However, these applications were only adopted for a specific situation in an isolated manner. Thus, a user-friendly and integrated system has still to emerge that can incorporate a variety of computer-aided tools to facilitate decision-making by model users.

Knowledge-Based Systems (KBSs)
KBSs are considered suitable for solving problems that demand considerable expertise, judgment or rules of thumb. The system is often compiled and encrypted to create a run-only system, which is preferably installed on a microcomputer for office use. Areas of early applications of KBS technology include medical diagnosis, mineral exploration and chemical spectroscopy. In recent years, KBSs have been applied to simulate domain problems in a variety of fields [13-21].

Artificial neural networks (ANNs)
ANNs are based on our present understanding of the brain and its associated nervous systems [22]. They use processing elements connected by links of variable weights to form black box representations of systems. A typical ANN comprises several layers of interconnected neurons, each of which is connected to other neurons in the ensuing layer. Data are presented to the ANN via an input layer whilst an output layer holds the response of the network to the input. Some hidden layers may exist between the input layer and the output layer. All hidden and output neurons process their inputs by multiplying each input by its weight, summing the product, and then processing the sum using a nonlinear transfer function to generate a result. The data-driven models have the ability to learn complex model functions from examples.

Fuzzy Inference Systems
Fuzzy logic is particularly useful in modeling complex and imprecise behaviors [23]. Under the fuzzy set theory, elements of a fuzzy set are mapped to a universe of membership values using a function-theoretic form belonging to the close interval from 0 to 1. This step is analogous to the estimation of probability in stochastic models. Moreover, the fuzzy logic based modeling operates on an ‘if-then’ principle. Both the vector of fuzzy explanatory variables or premises in the ‘if’ clause and the consequence in the ‘then’ statement are in the form of fuzzy set with membership functions.

The Prototype System

Literature review has been undertaken to identify the most appropriate recent methodologies and algorithms involving the integrated KBS on flow and water quality in order to improve the usability and display of numerical models. It should be noted that the user can always overrule any options and recommendations provided by the system. However, a mechanism is built-in to ensure that the user’s overruled input is reasonable and consistent. In other words, it plays the role of a knowledgeable assistant only. In order to facilitate the development of the KBS, an expert system programming environment or shell available in the market will be employed. With rapid developments in recent years, many successful expert system shells are now available. For this study, the requirements are the latest microcomputer-based shell that combines KBS technology with object-oriented programming, a relational database, graphics capabilities and debugging tools under an internet client/server network.

Expert System Shell
The development environment adopted for this prototype integrated system is VISUAL RULE STUDIO (VRS), which acts as an ActiveX Designer under the Microsoft Visual Basic

programming environment [24]. It incorporates a variety of knowledge representation schemes, different inference mechanisms and capabilities to interface with external programs in the windows environment. VRS provides an interactive windows-based user interface that runs under the conventions of Microsoft Windows. Under this system, any types of display windows, such as form, checkbox group, list box, command button, textbox, option button, picture box, etc., can be represented as objects, and each of them possesses their own special properties.

System Architecture
Figure 1 shows the system architecture of the prototype system. Besides the usual components in a typical KBS, namely, knowledge base, inference mechanism, session context, user interface, knowledge acquisition and explanation modules, it also incorporates an ANN tool, fuzzy rule system, and a database.

Knowledge acquisition and representation
The domain knowledge used has been acquired from written documents such as journal articles, textbooks and manuals, and complemented by interviews with experts. In acquiring knowledge, it is found that the outcomes are better to work together with the expert in the context of solving particular actual problems, instead of directly posing questions about rules. In order to tailor for each type of domain knowledge in the knowledge base and to take advantages of the characteristics of each method, hybrid knowledge representation schemes, including object-oriented programming, procedural methods, and production rules are employed to express engineering heuristics and standard design knowledge.

Object-oriented programming
Figure 2 shows the details of the blackboard architecture, which are classified into knowledge modules and the blackboard. Knowledge modules corresponding to procedural expertise knowledge is divided into Decision Process and Process Control whilst objects in the blackboard are basically classified into Decision Stage and Decision Entities. The blackboard is partitioned into a number of hierarchical levels, corresponding to different stages of the decision process. Decision Stage only comprises a single object whereas there are several objects in the Decision Entities level. Data inside Decision Stage are employed by the Process Control knowledge modules to determine the next possible action, or to check the validity of the function triggered by the user. Forward chaining inference mechanism is employed here to derive the next process. After a specific decision stage has been satisfied, the pertinent Decision Stage indicator will be assigned one of the preset values.

Process Control modules ensure the proper and effective application of knowledge in Decision Process modules and undertake conflict resolution. They evaluate the current attribute values in Decision Stage of the blackboard, which provides the indicator to assist this decision making. All primary tasks in Process Control module are expressed on command buttons together with procedural methods attached. Process Control knowledge modules work closely with the user-interface module to produce user-friendly main menu displays. The Main Decision Process class monitors the decision stage of all key tasks during the decision process and decides either to continue to next step or to prompt a warning message. Moreover, the relevant entries and decision parameters under Decision Entities, the corresponding attribute values of Decision Stage are synchronized through the Process Control knowledge modules. The procedural method attached to Decision Process modules is processed when the value of the attribute changes, either by assignment under another method or by the user. A mixed problem-solving strategy is adopted. After the user supply the relevant data during each decision stage, the system will determine the execution order of different decision knowledge modules.

Procedural methods
Procedural knowledge expressed using methods are often represented as program codes attached to attributes. Here, methods are often attached to the command buttons or option buttons. However, procedural methods are only attached to attributes of objects in Decision Process and Process Control knowledge modules representing the decision processes whilst objects in the blackboard representing the decision context do not embed any procedures.

Production rules
Some heuristic knowledge is represented in the IF/THEN/ELSE production rules with confidence factors that can be assigned either automatically, or in response to the user's request. These rules are a formal way of specifying how an expert selects and manipulate the model and its associated parameters. The following is a typical example of the production rules in this prototype system.

RULE to determine model dimensions: 4 of 5
IF the water depth is deep AND
    the density stratification in the vertical direction is significant
THEN the 3-dimensional numerical model is selected CF 75

In the production rules, the confidence factor (CF), with range is basically from 0 to 100, is employed as the determining factor to control the inference process for the evaluation of each parameter. The confidence factors, representing the degree of confidence of the statement, are set by weighted opinions of various experts based on their heuristic and experience. Appropriate rules will be executed depending on the responses of the user.

Incorporation of fuzzy logic
Fuzzy descriptions are also incorporated in production rules, with the membership functions set by weighted opinions of various experts. The user can specify continuously varying conditions, such as the pollutant source, with either fuzzy description or definite numerical values. The system can automatically transfer the numerical values into fuzzy description with the fuzzy membership curve to calculate its relevant confidence of membership before searching the rule base. Figure 3 shows the fuzzy description of the water depth with different curves representing the definitions of "very low", "low", "medium", "high" and "very high", respectively.

Incorporation of ANN
An ANN is used as the learning mechanism to transfer engineering experience into knowledge in the calibration of model parameters. The user has to choose appropriate answers to questions relating to the physical environment, data availability, demand on accuracy, demand on efficiency, and so on. The integrated system then matches the selected answers with each rule, executes the appropriate ones, and confirms their priority order on the basis of an ANN. A back-propagation architecture with 36 inputs (one attribute and one confidence factor for each parameter), 5 outputs (water depth, velocities and water quality at specific point of interest), and one hidden layer of 18 nodes is created for the problem. The control parameters include a learning rate of 1.0 and a momentum factor of 0.5. The initial network weights are assumed with uniformly distributed random values between -0.5 to 0.5. A linear function is employed for the input layer whilst the S-shaped sigmoid curve is used as a transfer function on each neuron to represent the input-output relation in the hidden layer

and output layer.

In order to evaluate the training performance, the normalized root-mean-square error (NRMSE) between target and output results is computed. Let N be the number of testing example, T_{ij}, O_{ij} be the target values and the computed value of the i^{th} test example and j^{th} output node respectively, and $\overline{T_j}$ be the average target value of j^{th} output node, then the definition of the above-mentioned statistical quantity is as follows:

$$NRMSE = \frac{\sum_{i=1}^{N}\sum_{j=1}^{4}(T_{ij} - O_{ij})^2}{\sum_{i=1}^{N}\sum_{j=1}^{4}(T_{ij} - \overline{T_j})^2} \qquad (1)$$

Figure 4 shows the relationship between NRMSE and number of training cycles. After about 30 cycles, self-learning mechanism is accomplished by the use of ANN. System validation is performed through the validation process of the ANN and by comparison of the results with those by the experts. It should be noted that the knowledge base is dynamic and if more input-output data pairs are provided by other experts in different locations of the world, better output results will be acquired via the generalization capability of the ANN.

Inference engine

The inference engine controls the strategies in the selection of procedure methods and production rules from the knowledge base to derive a conclusion or decision context. A hybrid reasoning strategy that combines forward and backward reasoning schemes is adopted. Given the current input values possibly with some unknowns initially, forward chaining inference is employed to infer the output values. Finally, when no unknown input can alter the current decision significantly, the prototype arrives at a conclusion. Backward chaining reasoning is employed, however, if the system has not arrived at a certain conclusion with a defined threshold value of confidence factor. The system will then automatically highlight the unknown input units that have significantly effect on the current most plausible conclusion, and prompt the user to enter their values. In this way, it is able to arrive at a reasonable conclusion with minimum information.

Moreover, it is tailored that all the decision steps are seen explicitly on the main screen display. The validity of the user's choice on the preferred sequence of decision processes is checked by Process Control knowledge modules, which act opportunistically upon being triggered. Moreover, an event-driven inference processing mechanism is adopted so that the ensuing action of the system will depend on the input made by the user. For example, if a 3-dimensional model has been selected, the user is prompted to enter the vertical spatial grid size. However, this question will not be prompted in case a 2-dimensional model has been chosen.

User-friendly interface

The system offers a friendly user interface. Whilst input data entries are kept at minimum, they are provided by the user mostly through selection of appropriate values of parameters from the menus and answers to the queries made by the system. If the input data provided by the user is not within the specified range, it will be rejected and warning message will be prompted.

Verification and Validation

Verification and validation of the prototype system are undertaken by applying to several real problems in Hong Kong coastal waters. The application case studies involve the establishment of several numerical models on coastal flow and water quality in Hong Kong, which encapsulate a few strategically chosen locations such as Pearl River Estuary [25-26], Shing Mun River [27], and Tolo Harbour [28-30]. Each of these water bodies is subject to a distinct environmental problem. The areas around the Pearl River Estuary, which is the largest river system in the South China, have been prospering during the past decade in such a dramatic rate that almost all the resources of the estuary are involved. This results in a worsening and deteriorating water quality as evidenced by the increasing occurrence of algal blooms recently. Shing Mun River is one of the most representative river channels in Hong Kong. The water quality of Tolo Harbour, which is an almost land-locked embayment with merely a narrow outlet channel, has declined drastically in recent years. It exhibits high algal growth and is in particular highly eutrophic. The details and modeling results for these areas, which can be found in the above references, are not reiterated here. This prototype system is applied to these modeling case studies so as to assist the selection of appropriate model and associated parameters and to evaluate the outcomes of employing a variety of methods to manipulate either accuracy or effectiveness. Figure 5 shows an example screen displaying user-friendly model selection for the application example.

Several numerical experiments were then conducted on these study areas to investigate the effects of applying different methods to manipulate either accuracy and/or efficiency. For each trial, only one design parameter was varied in order to study its effect. The improvement in accuracy was measured by comparing the NMSE for the two cases with and without the specified method applied, expressed in a percentage format. The improvement in efficiency was measured by comparing the required computer processing unit time for the two cases, also expressed in a percentage format. Table 1 shows a summary of comparisons in selected experiments regarding accuracy and/or efficiency improvement. In applying the integrated system to simulate case studies in Hong Kong coastal areas, it was found that the processes used in hydro-informatics and manipulation direction, were reasonable, since similar manipulation processes are also accomplished by human expert counterparts. In all the case studies, the most appropriate selection and manipulation of numerical model was effective, with regard to: bathymetry; data availability; demand on accuracy; demand on efficiency; and, numerical stability.

Conclusions

An integrated prototype KBS, which assists in making decisions on modeling process of flow and water quality, was developed and implemented. It incorporates an ANN for training of water quality parameters and a fuzzy rule base for representation of the heuristic knowledge. It is shown that the hybrid application of these latest AI technologies is appropriate to act as repository for heuristic knowledge. The knowledge base is transparent and can easily be updated, which render the prototype KBS an ideal tool for incremental programming. The prototype system has been successfully calibrated in Hong Kong conditions. This prototype system, serving both as a design aid as well as a training tool, is able to allow hydraulic engineers and environmental engineers to become acquainted with up-to-date flow and water quality simulation tools. Moreover, the system can quickly assist policy-makers in reaching decisions and also furnish convenient and open information service on water quality for the general public.

Acknowledgements

This research was supported by the Research Grants Council of Hong Kong (PolyU5132/04E).

References

[1] Leendertse, J.J., *Aspects of a computational model for long-period water- wave propagation*. RAND Corporation, Memo. RM-5294-PR, 1967.
[2] Abbott, M.B., "Review of recent developments in coastal modelling". In Falconer, R.A., Goodwin, P. and Mathew, R.G. S. (eds.), *Proceedings of International Conference on Hydraulic and Environmental Modelling of Coastal, Estuarine, and River Waters*. Aldershot, 1986.
[3] Abbott, M.B., *Hydroinformatics: Information Technology and the Aquatic Environment*. Avebury Technical, Aldershot, Brookfield, 1991.
[4] Adeli, H. and Hung, Shih-Lin, *Machine Learning: Neural Networks, Genetic Algorithms, and Fuzzy Systems*. Wiley, New York, 1995.
[5] Chau, K.W. and Chen, W., "An example of expert system on numerical modelling system in coastal processes", *Advances in Engineering Software*, 32(9), 2001, 695-703.
[6] Zou, R., Lung, W.S. and Guo, H.C., "Neural network embedded Monte Carlo approach for water quality modeling under input information uncertainty", *Journal of Computing in Civil Engineering, ASCE*, 16(2), 2002, 135-142.
[7] Chau, K.W., "River stage forecasting with particle swarm optimization," *Lecture Notes in Computer Science*, 3029, 2004, 1166-1173.
[8] Chen, Q.W. and Mynett, A.E., "Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake", *Ecological Modelling*, 162 (1-2), 2003, 55-67.
[9] Cheng, C.T. and Chau, K.W., "Fuzzy iteration methodology for reservoir flood control operation," *Journal of the American Water Resources Association*, 37(5), 2001, 1381-1388.
[10] Mulligan, A.E. and Brown, L.C., "Genetic algorithms for calibrating water quality models", *Journal of Environmental Engineering, ASCE*, 124(3), 1998, 202-211.
[11] Chau, K.W., "Calibration of flow and water quality modeling using genetic algorithm," *Lecture Notes in Artificial Intelligence*, 2557, 2002, 720-720.
[12] Chen, L. "A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data," *International Journal of Remote Sensing*, 24(11), 2003, 2265-2275.
[13] Chau, K.W. "An expert system for the design of gravity-type vertical seawalls," *Engineering Applications of Artificial Intelligence*, 5(4), 1992, 363-367.
[14] Chau, K.W. and Yang, W.W. "Development of an integrated expert system for fluvial hydrodynamics," *Advances in Engineering Software*, 17(3), 1993, 165-172.
[15] Chau, K.W. and Zhang, X.N. "An expert system for flow routing in a river network," *Advances in Engineering Software*, 22(3), 1995, 139-146.
[16] Chau, K.W. and Ng, V. "A knowledge-based expert system for design of thrust blocks for water pipelines in Hong Kong," *Journal of Water Supply Research and Technology – Aqua*, 45(2), 1996, 96-99.
[17] Chau, K.W. and Albermani, F. "Expert system application on preliminary design of

liquid retaining structures," *Expert Systems with Applications*, 22(2), 2002, 169-178.
[18] Chau, K.W. and Anson, M. "A knowledge-based system for construction site level facilities layout," *Lecture Notes in Artificial Intelligence*, 2358, 2002, 393-402.
[19] Chau, K.W. and Albermani, F., "A knowledge-based system for liquid retaining structure design with blackboard architecture," *Building and Environment*, 40(1), 2005, 73-81.
[20] Lin, S. and Albermani, F. "Lattice-dome design using a knowledge-based system approach," *Computer-aided Civil and Infrastructure Engineering*, 16(4), 2001, 268-286.
[21] Ranga Rao, A.V. and Sundaravadivelu, R. "A knowledge based expert system for design of berthing structures," *Ocean Engineering*, 26(7), 1999, 653-673.
[22] Rumelhart, D.E., Widrow, B. and Lehr, M.A. "The basic ideas in neural networks," *Communications of the ACM*, 37(3), 1994, 87-92.
[23] Zadeh, L. and Kacprzyk, J., Fuzzy Logic for the Management of Uncertainty. John Wiley, New York, 1992.
[24] Machine Corporation Corporation, *User's Guide for Visual Rule Studio*. Rule Machines Corporation, Indialantic, 1998.
[25] Chau, K.W. and Jiang, Y.W. "3D numerical model for Pearl River Estuary," *Journal of Hydraulic Engineering, ASCE*, 127(1), 2001, 72-82.
[26] Chau, K.W. and Jiang, Y.W. "Three-dimensional pollutant transport model for the Pearl River Estuary," *Water Research*, 36(8), 2002, 2029-2039.
[27] Chau, K.W. and Lee, J.H.W. "Mathematical modelling of Shing Mun River network," *Advances in Water Resources*, 14(3), 1991, 106-112.
[28] Chau, K.W. and Jin, H.S. "Numerical solution of two-layer, two-dimensional tidal flow in a boundary fitted orthogonal curvilinear coordinate system," *International Journal for Numerical Methods in Fluids*, 21(11), 1995, 1087-1107.
[29] Chau, K.W. and Jin, H.S. "Eutrophication model for a coastal bay in Hong Kong," *Journal of Environmental Engineering, ASCE*, 124(7), 1998, 628-638.
[30] Chau, K.W., Jin, H.S. and Sin, Y.S. "A finite difference model of 2-d tidal flow in Tolo Harbour, Hong Kong," *Applied Mathematical Modelling*, 20(4), 1996, 321-328.

Table 1. Examples of comparison of accuracy or efficiency improvement from this case study

Change in scheme/algorithm	Improvement in accuracy	Improvement in efficiency
Use of 1-dimension instead of 2-dimension	-11%	50%
Use of 2-dimension instead of 3-dimension	-10%	60%
Incorporation of major tidal constituent only	-7%	8%
Assuming constant friction coefficient	-6%	12%
Increase in time step size	-6%	30%
Use of minimum number of variables	-5%	10%
Increase in horizontal grid spacing	-5%	35%
Use of lower order scheme	-4%	25%
Use of explicit scheme instead of implicit scheme	-3%	20%

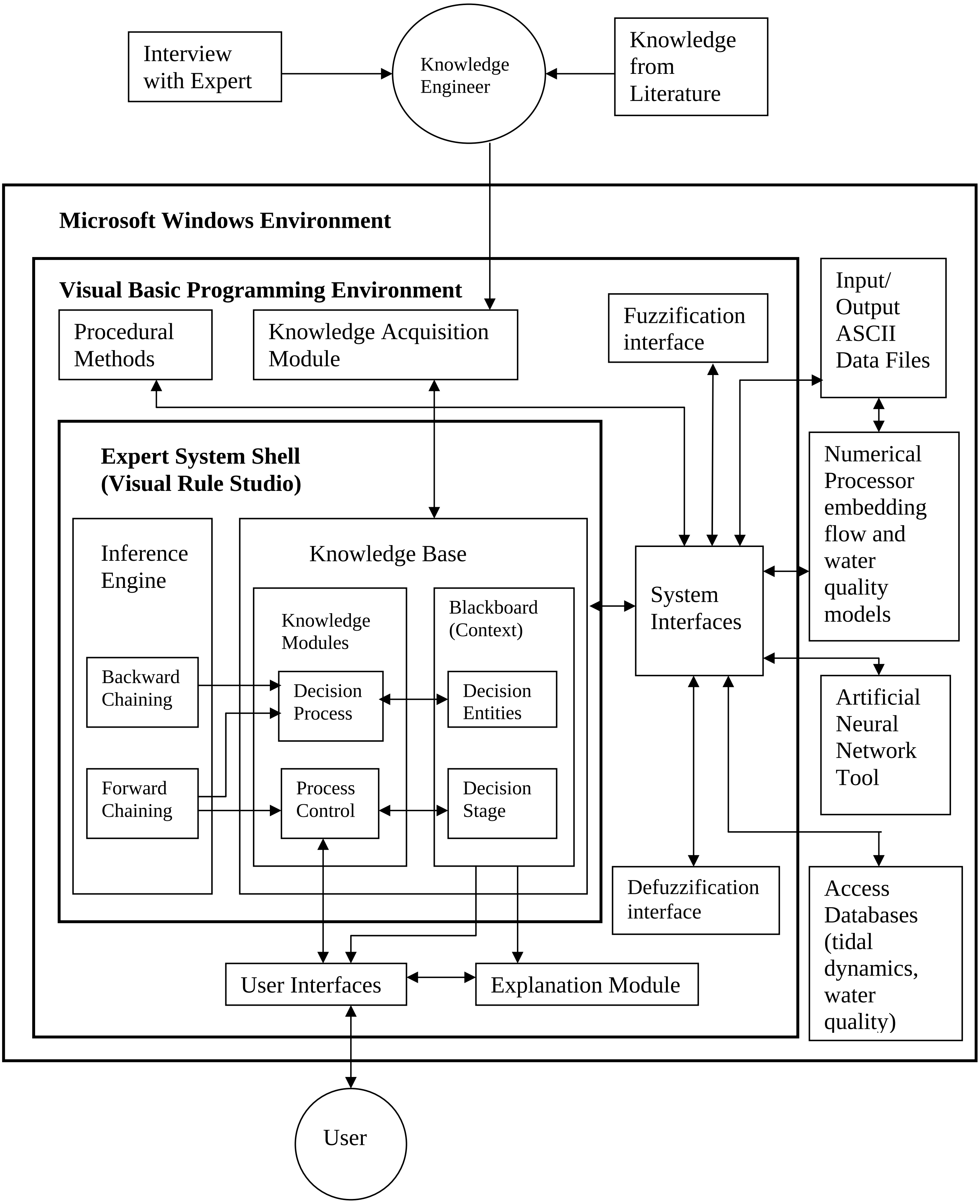

Figure 1. Schematic diagram of the prototype integrated system

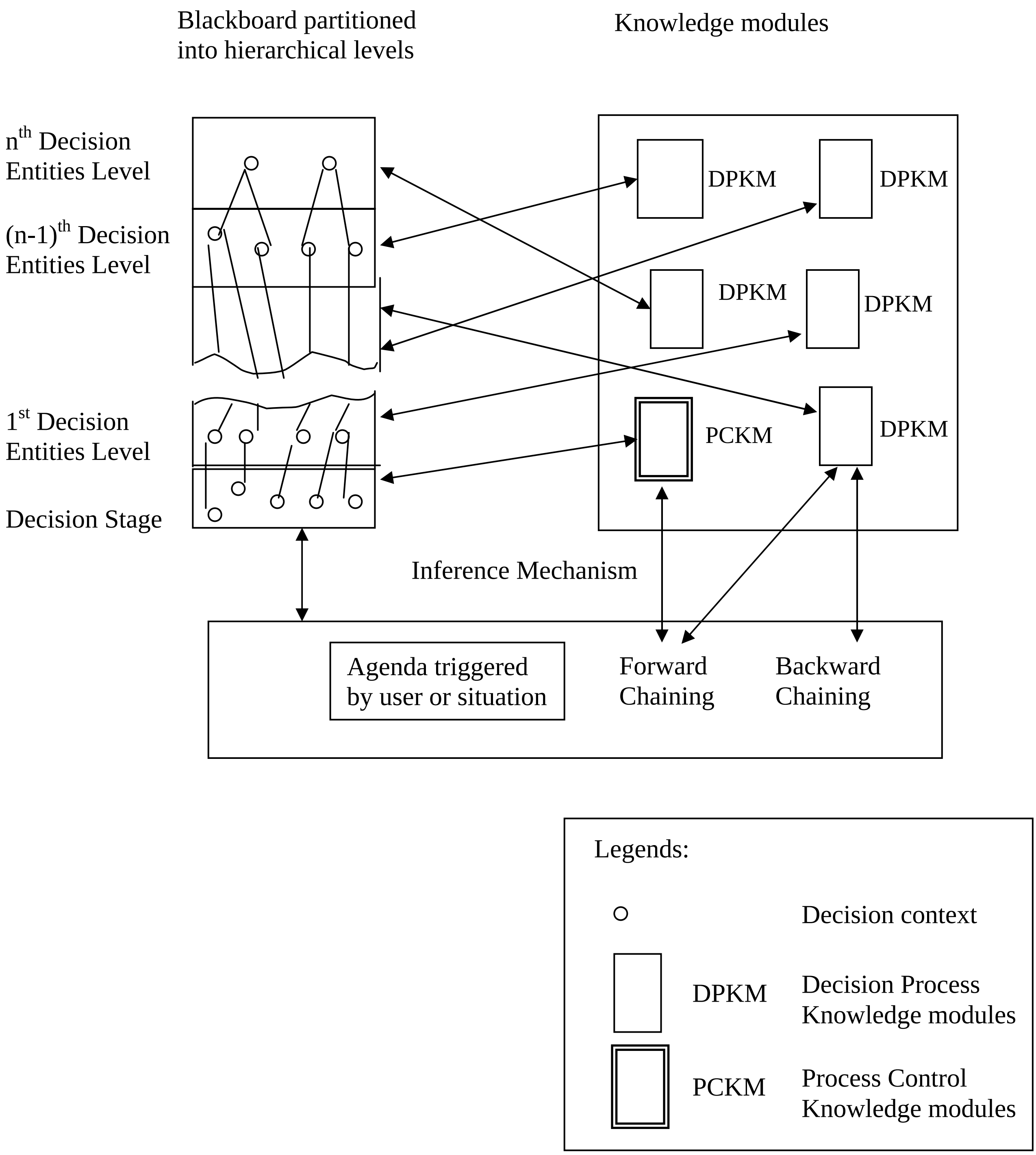

Figure 2. Details of the blackboard architecture

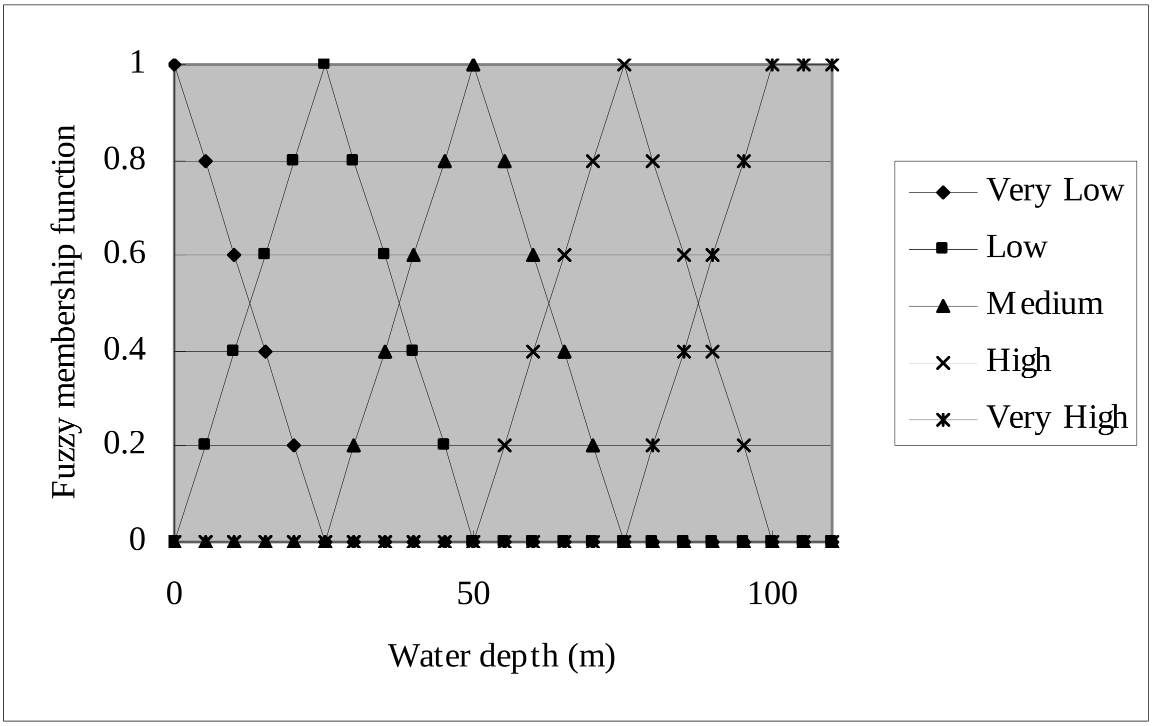

Figure 3. Fuzzy membership function of water depth

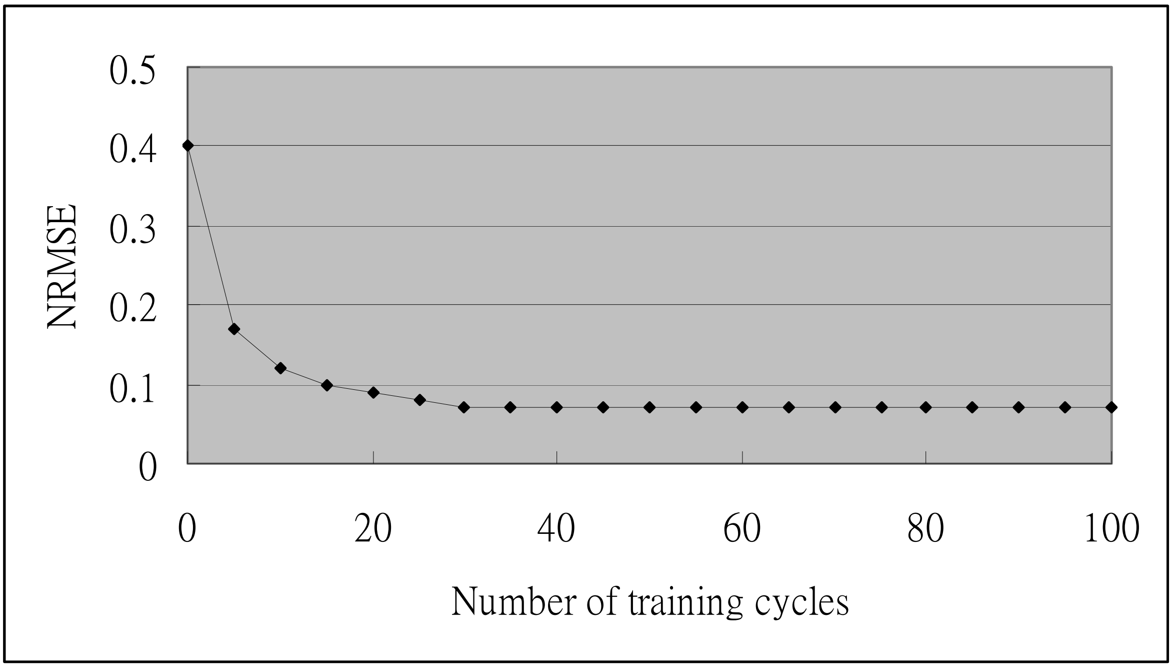

Figure 4. Relationship between NRMSE and number of training cycles

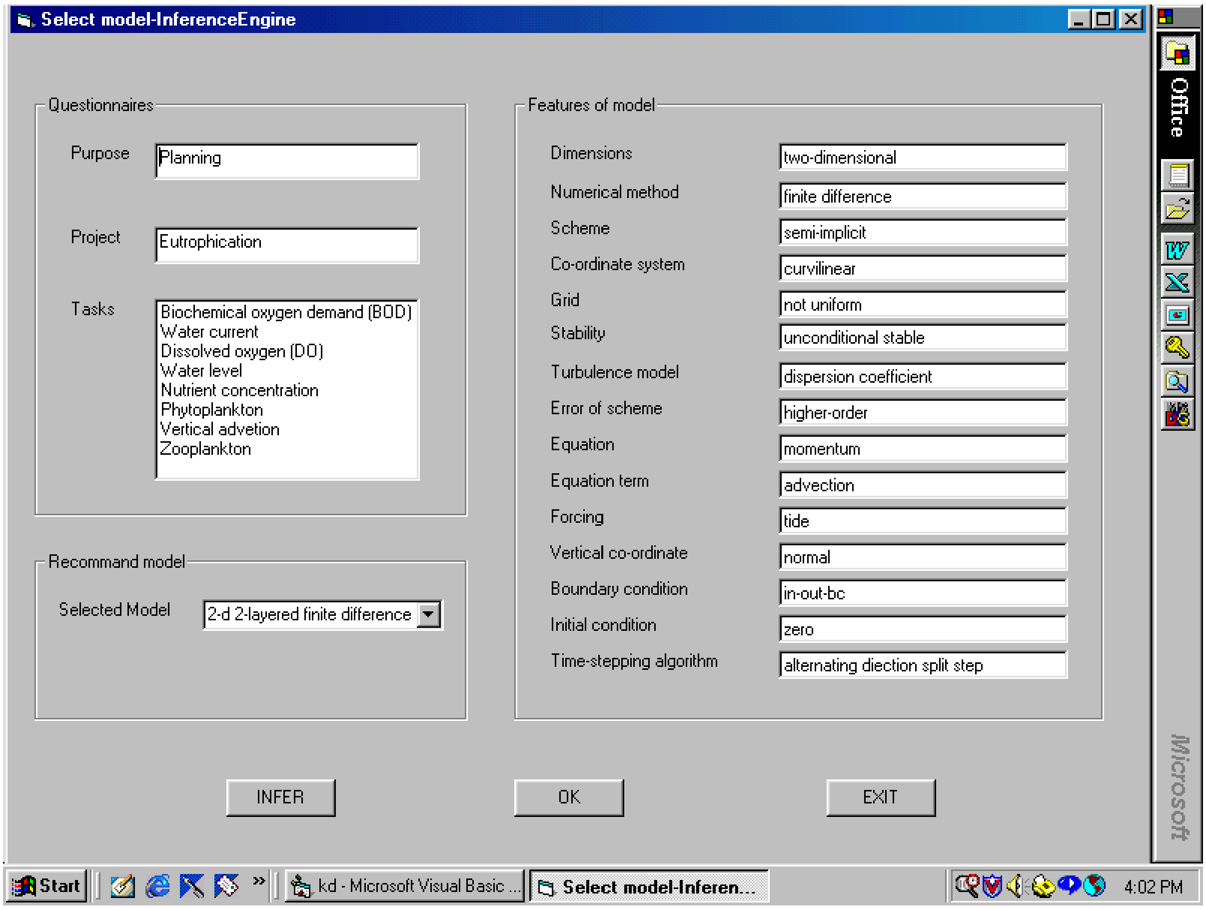

Figure 5. Screen displaying user-friendly model selection for the application example