Intelligent Manipulation of Calibration Parameters in Numerical Modeling

K.W. Chau

Department of Civil & Structural Engineering, Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong

Abstract

Numerical models are often employed to simulate flow and water quality processes in coastal areas. Model manipulation is always required, particularly during the set-up, since a slight change of model parameters may lead to quite different results. Nowadays, with the maturity of numerical modeling paradigms, there is abundant heuristic knowledge relating design parameters to outcomes. Techniques are also available to improve modeling accuracy and/or speed. It is desirable to incorporate the existing heuristic knowledge about model manipulation in the form of a rule base and to furnish intelligent manipulation of calibration parameters. In this paper, a prototype expert system for model manipulation for flow and water quality is developed and implemented by employing an expert system shell. The architecture and main components of the system are delineated. Expertise in model manipulation, including knowledge as to what is the correct manipulation direction and how to enhance manipulation effectiveness, are incorporated. It is shown that, through the successful development of this prototype system, expert system technology can be integrated into numerical modeling systems.

Introduction

With the advancement of computer technology, numerical models are often employed to simulate various flow and water quality processes in coastal areas (Blumberg and Mellor, 1987; Chau and Jin, 1998; Chau and Lee, 1991; Cheng et al. 1984; Leendertse, 1967). Conventionally, the emphasis has been placed on algorithmic procedures to solve specific coastal problems. These numerical models, being insufficiently user-friendly, lack knowledge transfers in model interpretation. This results in significant constraints on model uses and large gaps between model developers and practitioners. It is usual that, especially for non-expert users, the length of procedures for model manipulation depends largely on their experience.

Model manipulation is always required, particularly during the set-up of the model, since a slight change of the parameters may lead to quite different results. The conventional procedure of model manipulation can be described as: roughly select a numerical model with moderate confidence of correctness, execute it, estimate the correctness of the model results, determine the correct direction to modify certain model parameters, then revise the parameters and improve the pertinent confidence of correctness. This process is iterated for several cycles until the results meet the threshold of mathematical and physical correctness preset by the user. The procedure is considered completed when the simulation of real phenomena becomes satisfactory. Selection and manipulation of an appropriate numerical model to solve a practical problem is a specialized task, entailing detailed knowledge of the applications and limitations of the model. Many model users do not possess the requisite knowledge to glean their input data, build algorithmic models and evaluate their results. This may produce inferior design and cause under-utilization, or even total failure, of these

models. Experienced modelers can determine a model failure based on the comparison of the simulated results with real data as well as a heuristic judgement of the key environmental behavior. It is difficult for a novice user to ascertain whether or not his/her modeling efforts have been successful. Moreover, measurements for comparison are often absent.

Previous efforts have been devoted to accommodating a much wider range of end-users, through a menu of parameter specification, automatic grid formation, pre-processing and post-processing facilities, etc. These tools act as intelligent front-ends to support the handling of simulation models for specific hydrological or water quality problems (Abbott, 1991; Knight and Petridis, 1992). Yet they do not address the core problem of knowledge elicitation and transfer. Ragas et al. (1997) suggested, but did not implement, the incorporation of an expert system into modeling, in order to deal with uncertainty in model predictions, after they had compared a number of UK and USA models. An expert system approach has been adopted for river planning. Chau and Yang (1993) implemented an integrated expert system for fluvial hydrodynamics. Jamieson and Fedra (1996) developed a decision-support system for efficient river basin planning and management. Ghosh Bobba et al. (2000) applied environmental models through an intelligent system to different hydrological systems. These works are, however, limited to one-dimensional modeling systems, and represent only a minute portion of knowledge in this field. Moreover, their knowledge bases include heuristic rules for model selection but not for model manipulation.

Nowadays, with the maturity of modeling paradigms, there is abundant heuristic knowledge relating design parameters to outcomes. A variety of techniques is also available to improve modeling accuracy and/or speed. As such, it is desirable to incorporate the existing heuristic knowledge of model manipulation in the form of a rule base and to furnish intelligent manipulation of calibration parameters. In this paper, a prototype expert system for model manipulation for flow and water quality is developed and implemented by employing an expert system shell.

Model Manipulation

Abbott (1993) described numerical modeling as a procedure that transformed our knowledge of natural water phenomena into digital forms, let the computer run for the simulation, and translated the numerical results back into a comprehensible knowledge format. It is through this process that our understanding of nature can be enhanced. The processing from knowledge to information is a procedure of selection of a suitable model together with model parameters whilst the processing from information to knowledge represents post-processing and calibration of data from the results. Model manipulation is a mixed procedure of feed back and modification comprising the two processes described. A great deal of knowledge of model manipulation exists. This includes knowledge of real physical observations, the mathematical description of water movement or water quality, the discretization of governing equations for physical and chemical processes, schemes to solve the discretized equations effectively and accurately, and output analysis. For an experienced modeler, the knowledge mentioned above may be used unconsciously.

The ultimate goal of model manipulation in coastal engineering is to acquire satisfactory simulation. However, since a computer is used and the memory and speed of a computer are often limited, a balance should be struck between the modeling accuracy and speed. It is noticeable that modelers usually keep certain fundamental parameters unchanged during the manipulation process. For instance, when Leendertse (1967) reported that the Alternate

Direction Implicit (ADI) model could be applied in two-dimensional flow modeling, both the discretization method and the algorithm scheme were kept unchanged by others. When researchers were used to two-dimensional coastal modeling, they varied only the bottom friction coefficient (Chau and Jin, 1995). In water quality modeling, Baird and Whitelaw (1992) reported that the algal behavior was related intimately to both its respiration rate and the water temperature. Model users will consider sunlight intensity variation within the water column when simulating the eutrophication phenomenon (Chau and Jin, 1998). These examples reflect that human intelligence uses existing knowledge to reduce the number of choices in order to raise the effectiveness of model manipulation. Each time, they tend to alter merely one or two parameters. This is because if they modify many parameters at the same time, they may easily get lost in the manipulation direction.

Expert System on Model Manipulation

Programming development environment

A hybrid expert system shell that couples the advantage of both production rules and the object-oriented programming paradigm is employed here. All the usual control objects, including textbox, option button, command button, check box, etc., are furnished by this programming environment. Another key deciding factor in adopting this shell is the user-friendliness of the interface. Moreover, this software has the ability to link or execute external programs coded in other traditional programming languages.

System Architecture

This system comprises many modules, with part of the integration being effected through add-in tools. Besides the usual components in a typical expert system, namely, knowledge base, inference mechanism, session context, user interface, knowledge acquisition facility and explanation module, it also incorporates a toolbox, executable numerical models and databases. The architecture of the prototype system is shown in Figure 1.

The knowledge base

A great deal of knowledge about model manipulation can be represented in the form of rules inside the knowledge base. By using the object-oriented expert system shell, rules can be written in the Production Rule Language form, and easily integrated into the modeling system. There are also many fuzzy processes to estimate whether or not modeling results meet the demands of the user as to modeling accuracy or speed. The knowledge base incorporates the whole set of inference rules relating to the manipulation direction and the user's requirements. The following example gives a typical production rule which incorporates the fuzzy description:

RULE to manipulate dimensions of model: 3 of 10
IF a 2-dimensional numerical model is currently used AND
        demand on accuracy is very high AND
        computed error is very large
THEN a 3-dimensional numerical model is selected with a confidence factor of 75.

The IF clause of the above rule statement describes the premises or conditions under which a 2-dimensional numerical model should be used: the demand for accuracy should be very high

and the computed error should be very large for the conclusion to be fulfilled and hence the rule to be fired. The THEN statement of the rule gives the conclusion that a 3-dimensional numerical model is selected with a confidence factor of 75. The confidence factor is employed as the determining factor to control the inference process for model manipulation for each possible parameter. Its range is basically from 0 to 100, representing the degree of confidence with which the statement is known. The confidence factors are set by experts based on heuristic and experience.

The inference mechanism

After the user has selected a model prototype and has specified his preference for model accuracy or speed, the inference engine can use the mixed strategy of backward chaining and forward chaining to give the suggested direction of manipulation. Forward chaining is employed to search from the user's responses to a query in order to modify the decision tree and to execute the numerical process. Backward chaining is used to find the direction of manipulation and to determine the queries and their possible responses, on the basis of production rules in the knowledge base. This inferencing process occurs in a cyclical manner until the preset threshold of error margin is met.

The user interface

The primary function of the user interface is to enable the user to specify the demand as to accuracy or speed, to describe the physical conditions of the simulated environment and to acquire output results from the design consultation. The user interface furnishes the requisite information to let the system make inferences from the rule base. Figure 2 shows a sample screen of the user interface for the selection of model and its associated parameters. The user interacts with the system by employing a natural language expression as far as possible, and by keeping the input data entry at a minimum. Communications with the system are directed mostly through the selection of appropriate values or parameters from menus and through replying answers to the system. Graphical user interfaces, consisting of layers of display screens and pop-up windows, are used for message transfer.

The numerical processor

The numerical processor, which can be executed to generate the numerical simulation of real phenomena, is the central component of traditional numerical modeling systems. These well-proven and validated models have often been developed in conventional languages. It is considered not cost-effective to replace them with ones whose development involves a great deal of work. Instead, these programs are written in embedded source codes. The integrated system is accompanied by a usage wizard, which provides assistance and guidance for use and direction for non-expert users.

This numerical processor is designed in object linking and embedding (OLE) automation format for use by other OLE-compliant applications if required. It is ready to run with various properties and methods of action. The selection of the modeling parameters is for setting the properties and methods of this numerical processor, which are listed in Tables 1and 2 respectively. Properties related to modeling are classified into six classes: dimensions, grid, scheme, turbulence, boundary conditions, and others. Methods related to result display include saving the data file, comparing with existing data information, monitoring the results under limitations, recording the images to show the result display, etc. It is designed to run

with a set of default values. This numerical processor is incorporated as a control object, and only the properties and methods are exposed to changes in model manipulation. The major advantage of this arrangement is that different modules of the modeling system exist independently and can be easily updated individually, thus facilitating the integration of expert system technology. Figure 3 shows the flowchart of the modeling manipulation process. Figure 4 shows a sample screen displaying the graphical user interface incorporated into numerical processing. Details, including the grid layout, a three-dimensional perspective view, a scaling legend, the tidal boundary conditions and certain key controlling parameters in a visual edit form, are shown during the execution of the numerical processor.

The toolbox

A primary function of the toolbox is to provide a comparison between the simulated results and the real observation data. The improvement in estimating modeling results depends on the technology of pattern recognition. The normalized root-mean-square error (NRMSE) between key field data and model results is computed to evaluate the performance of the model and its associated model parameters. It covers cases with a time series of data at a single point within the model domain, or instantaneous measurements at many locations, or a combination of both. Let N be the number of data location for comparison, n be the number of time interval in a time series of data for comparison, $T_{i,t}$, $O_{i,t}$ be the target values and the computed value of the ith data location and t^{th} time step respectively, and $\overline{T}$ be the average target value, then the definition of the above-mentioned statistical quantity is as follows:

$$NRMSE = \frac{\sum_{i=1}^{N}\sum_{t=1}^{n}(T_{i,t} - O_{i,t})^2}{\sum_{i=1}^{N}\sum_{t=1}^{n}(T_{i,t} - \overline{T})^2} \tag{1}$$

Figure 5 shows the membership functions for NRMSE, which represent the fuzzy logic of literal classification into very small, small, large, and very large. The toolbox also provides assistance for post-processing as well as pre-processing, such as drawing the boundary with the mouse, drawing the time variation curve of tidal boundary conditions, etc. The toolbox can be used to monitor model manipulation through prompting a warning message if the model results exceed the pre-set limitations. Automatic checks and warning prompts are designed to effect the verification between the results and the observation data.

Databases

A numerical model can only be verified by simulating a particular application on the basis of the correct description of a real situation. It is believed that more observation data would be useful to mimic the situation. Certain understandings about phenomena are in the form of data, whilst others are in the form of description statements. These data are concerned with dynamics (current, tidal elevation, or water depth), temperature and salinity, water quality (dissolved oxygen, phosphorus, or nitrogen), and their variations with time and location. The description statement can be written in the form of rules and included in the database. The prototype system integrates a large quantity of field data to describe real phenomena. The system is verified and validated by application to prototype problems in Hong Kong coastal waters, which are stored in databases. Details of the models can be found in previous literature (Chau and Lee, 1991; Chau and Jin, 1995; Chau and Jin, 1998; Chau and Jiang,

2001; Chau and Jiang, 2002) and are not reiterated here. Figure 6 shows a sample screen displaying an interactive graphical representation of the topography at the Pearl River Estuary.

Knowledge acquisition

The domain knowledge entailed in the development of this expert system is encoded mainly from a literature review and interviews with expert numerical modelers. Details pertinent to modeling application and selection have been reviewed to capture such knowledge or relationships in the form of rules. Currently no standard steps are generally recognized and model manipulation may vary from expert to expert. However, this prototype expert system can be further developed and updated through frequent usage and feedback from users. One advantage of an expert system approach is the transparency of rules and knowledge bases, which facilitates updating with new, additional knowledge.

Knowledge elicitation and representation

In this expert system, the same model manipulation as that carried out by its counterpart by a modeler is simulated. The direction and effectiveness of manipulation are two main factors in the process. Human intelligence about them can be extracted and described in the form of rules. After rules are incorporated into the code in a rule base, and an expert system technology is applied, the non-experienced user may benefit from the assistance provided by the rules in the knowledge base. The user can specify his preference for accuracy or speed, and a set of parameters would be suggested in the interface for modification.

When a model is chosen in the model selection procedure, certain parameters are kept unchanged, and only a few parameters are manipulated to improve the simulation. Under the terms of expert system technology, model templates are prepared as a kind of model frame. The frame consists of many slots, representing model parameters. Whilst certain slots cannot be changed, others can be modified to fit the specific environment. Thus the procedure becomes to seek suitable slots so that the model results coincide properly with real phenomena. In the modeling system, those parameters which can be modified are listed in the form of a parameter tree. Different choices, representing possible selections of branch, are available to that specific parameter. The tree is managed by production rules in the knowledge base. It has a full skeleton during the modeling selection. Some branches are cut off after the user has specified his demand for accuracy or speed. After model manipulation is completed, the tree skeleton becomes a solid tree with specified branches. Figure 7 shows an example of the tree formation for model manipulation.

Conclusions

Existing numerical models are insufficiently user-friendly and often result in significant constraints on their uses. As such, it is instrumental to incorporate the existing heuristic knowledge of model manipulation in the form of a rule base and to furnish intelligent manipulation of calibration parameters. This study transforms expert knowledge in model manipulation into a knowledge base, which is then integrated with numerical modeling. A prototype expert system for model manipulation for flow and water quality has been developed and implemented by employing a shell. Expertise in model manipulation, including the knowledge as to what is the correct manipulation direction and how to enhance manipulation effectiveness, are incorporated into the prototype system. It is shown that,

through the successful development of this prototype system, expert system technology can be integrated into numerical modeling systems. This can provide substantial assistance to novice users of these algorithmic models to determine whether or not digital sets generated by numerical modeling represent real phenomena. It is believed that further development of numerical modeling should be undertaken in this direction. In fact, the ideas presented in this paper are not only limited to coastal engineering but also applied to other settings.

Acknowledgments

References

1. Abbott, M.B. (1993), The electronic encapsulation of knowledge in hydraulics, hydrology and water resources, Advances in Water Resources 16(1), pp. 21-39
2. Baird, J.I. and Whitelaw, K. (1992), Water quality aspects of estuary modeling, Water Quality Modeling, Edited by Roger A. Falconer, pp. 119-126
3. Blumberg, A.F. and Mellor, G. (1987), A description of a three-dimensional coastal ocean circulation model. Three-dimensional coastal ocean models, ed. by N.S. Heaps, American Geophysical Union, Washington, D.C., pp. 1-16
4. Chau, K.W. and Jiang, Y.W. (2001), 3D numerical model for Pearl River Estuary, Journal of Hydraulic Engineering, ASCE, 127(1), pp. 72-82
5. Chau, K.W. and Jiang, Y.W. (2002), Three-dimensional pollutant transport model for the Pearl River Estuary, Water Research, 36(8), pp. 2029-2039
6. Chau, K.W. and Jin, H.S. (1995), 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), pp. 1087-1107
7. Chau, K.W. and Jin, H.S. (1998), Eutrophication model for a coastal bay in Hong Kong, Journal of Environmental Engineering, ASCE, 124(7), pp. 628-638
8. Chau, K.W. and Lee, J.H.W. (1991), Mathematical modelling of Shing Mun River network, Advances in Water Resources, 14(3), pp. 106-112
9. Chau, K.W. and Yang, W.W. (1993), Development of an integrated expert system for fluvial hydrodynamics, Advances in Engineering Software, 17(3), pp. 165-172
10. Cheng, R.T., Casulliv, V. and Milford, S.N. (1984), Eulerian-Lagrangian solution of the convection-dispersion equation in natural coordinates , Water Resources Research, 20(7), 944-952
11. Ghosh Bobba, A., Singh, V.P. and Bengtsson, L. (2000), Application of environmental models to different hydrological system, Ecological Modelling, 125(1), pp. 15-49
12. Jamieson, D.G. and Fedra, K. (1996), The 'WaterWare' decision-support system for river-basin planning. 1. Conceptual design, Journal of Hydrology, 177(3-4), pp. 163-175
13. Knight, B. and Petridis, M. (1992), Flowes: an intelligent computational fluid dynamics system, Engineering Applications of Artificial Intelligence, 5(1), pp. 51-58
14. Leendertse, J.J. (1967), Aspects of a computational model for long-period water- wave propagation. RAND Corporation, Memo. RM-5294-PR
15. Mellor, G.L. (1996), User's guide for a three-dimensional, primitive equation, numerical model, Princeton University Rep. Princeton University, Princeton, NJ
16. Ragas, A.M.J., Haans, J.L.M. and Leuven, R.S.E.W. (1997), Selecting water quality models for discharge permitting, European Water Pollution Control, 7(5), pp. 59-67

Table 1. Table listing primary properties of numerical processor

class	*Main properties*
Dimensions and Method	dimensions, numerical method, co-ordinate system, numerical method in vertical direction, vertical co-ordinate
Grid	grid shape, x-grid spacing, y-grid spacing, vertical grid spacing, grid uniform or not, type of point setting
Scheme	advection term, explicit or implicit, stability, error of scheme, alternating direction or not, algorithm
Turbulence	turbulence model, vertical eddy viscosity, horizontal eddy viscosity, vertical eddy diffusion, horizontal eddy diffusion, bottom drag coefficient
Boundary condition and driving force	wind at surface, tide at open boundary, river discharge at open boundary, value and variation at open boundary, value in close boundary, variation in close boundary, high order variation in boundary
Some others	Coriolis force, geometry, bathymetry, initial elevation conditions, initial current conditions, initial conditions of water quality variables

Table 2. Table listing methods of numerical processor

Method	*Description*
Initialization	Initialization of all data to become default values
Modify	Modify certain main properties or parameters
Save	Saving the data file
Run	Execution of numerical modeling
Pause	Temporary pause of execution of numerical modeling
Stop	End of execution of numerical modeling
Record	Recording images to display the results
Comparison	Comparing results with available field data for calibration
Warning	Prompting warning messages
Limitation	Monitoring the results under limitation
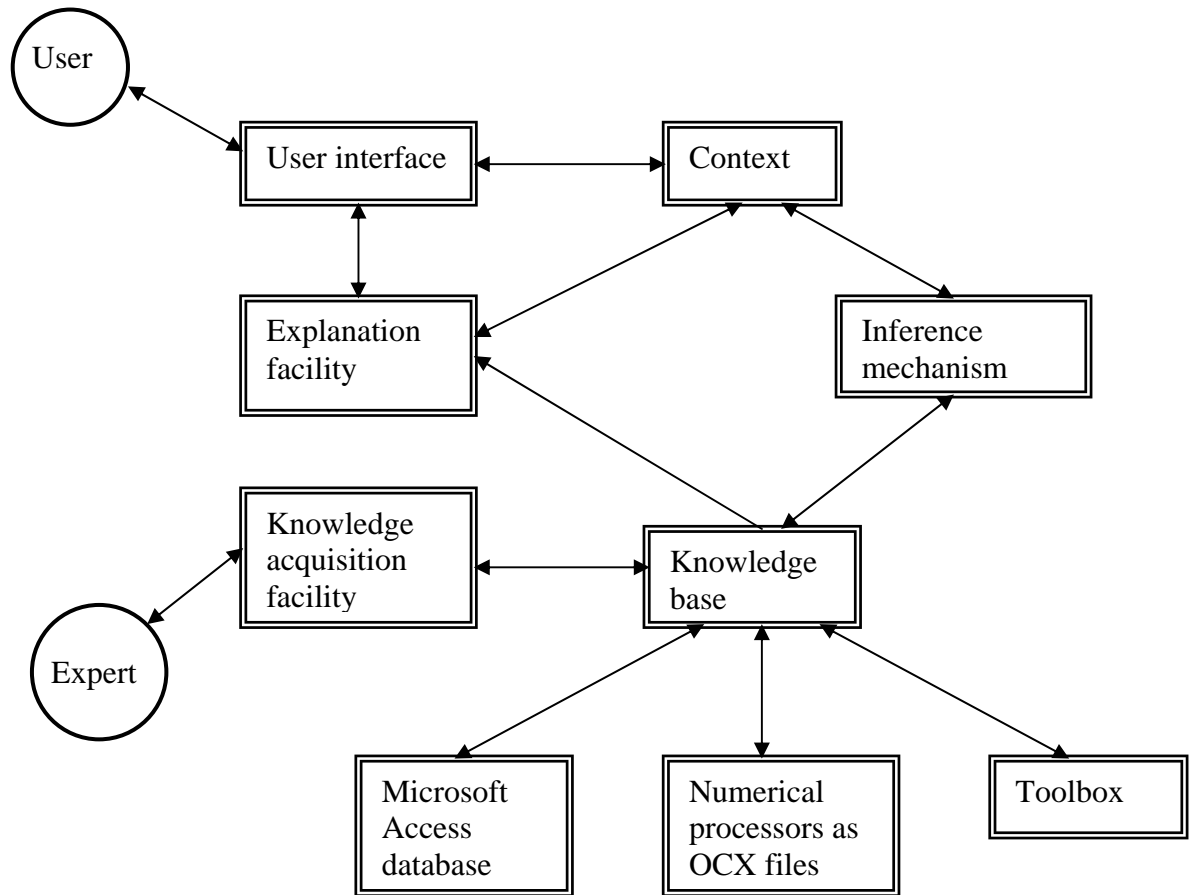
Figure 1 System architecture of the prototype system
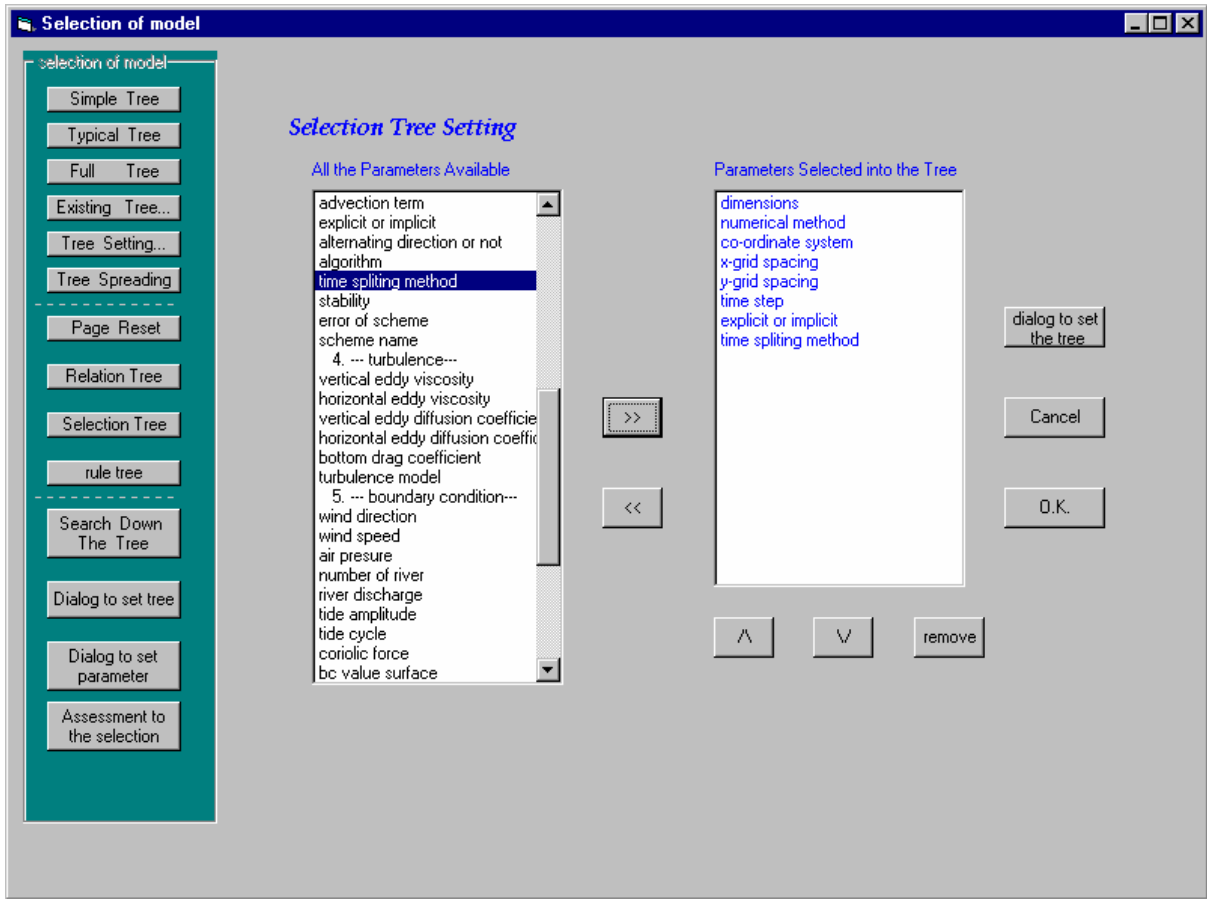
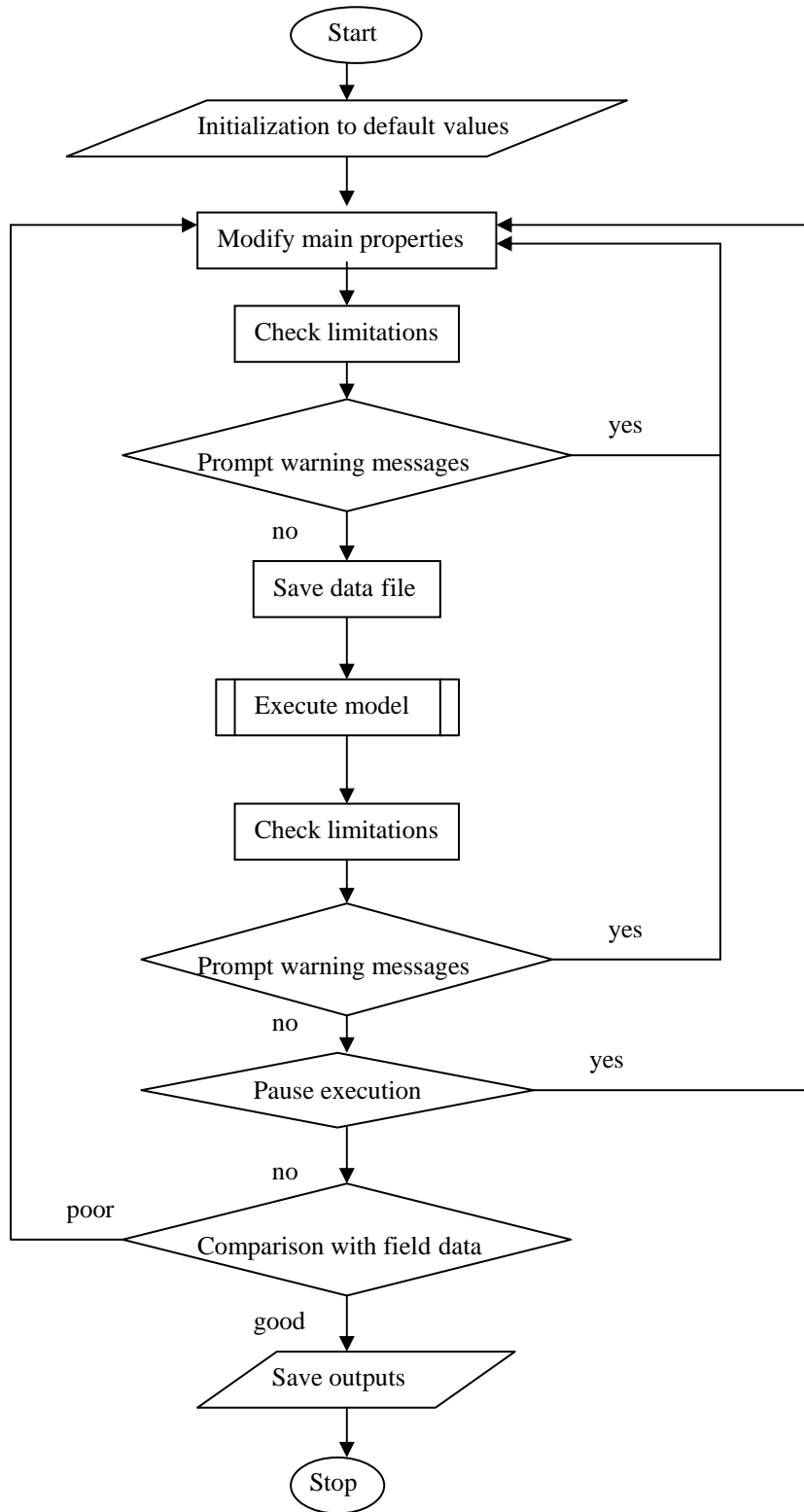Figure 2. A sample screen of the user interface for selection of model and its associated parameters

Start

Initialization to default values

Modify main properties

Check limitations

Prompt warning messages — yes

no

Save data file

Execute model

Check limitations

Prompt warning messages — yes

no

Pause execution — yes

no

Comparison with field data

poor

good

Save outputs

Stop

Figure 3. Flow chart showing the modeling manipulation process
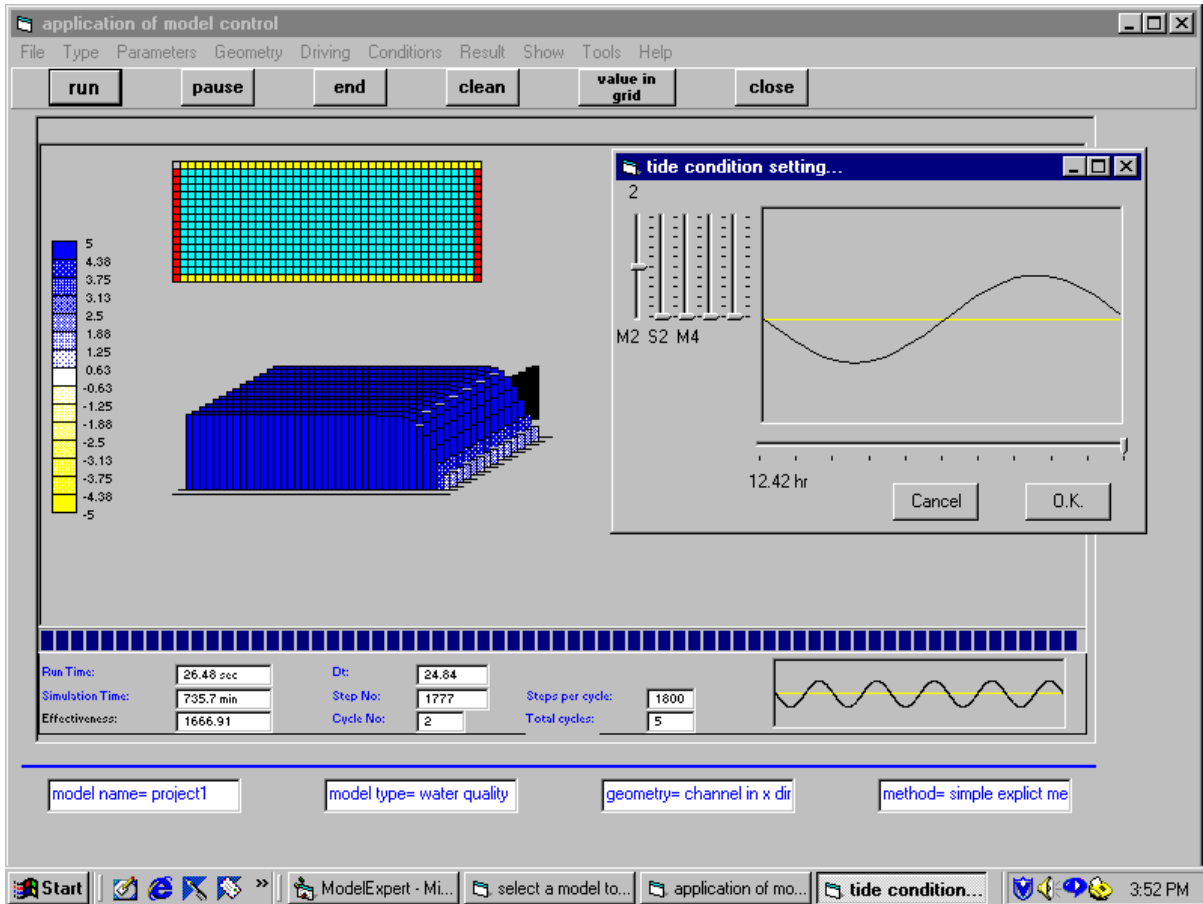
Figure 4. A sample screen displaying graphical user interface incorporated into numerical processing with details on grid layout, 3-D view, scaling legend, and tidal boundary condition
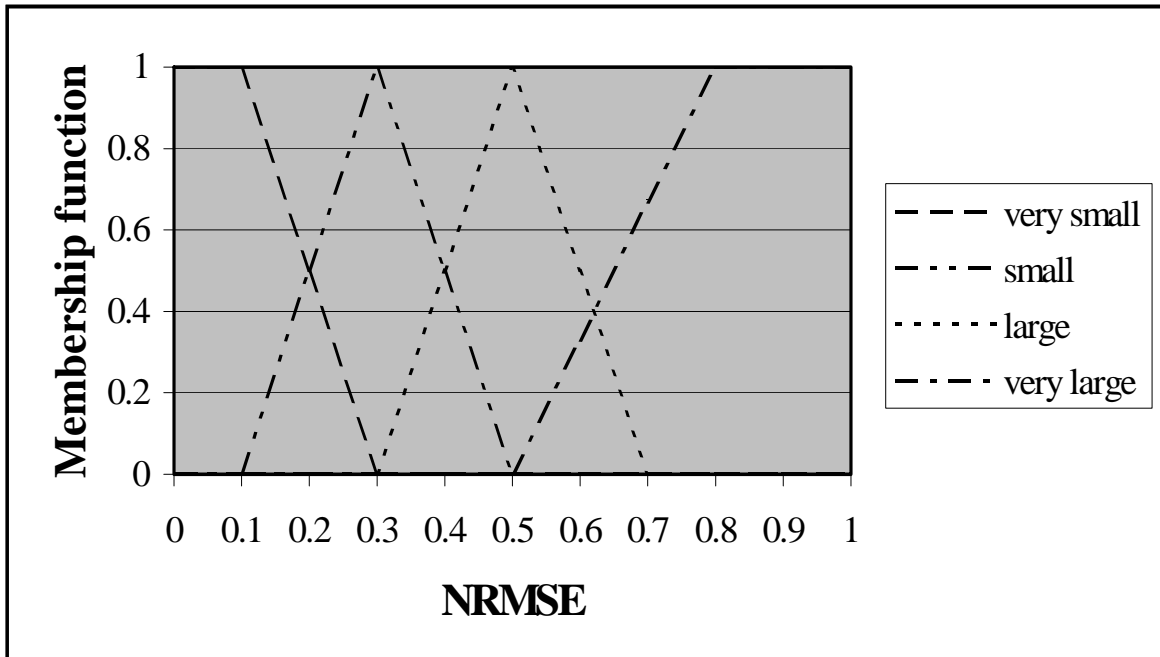
Figure 5. Membership functions for normalized root-mean-square error (NRMSE) representing the fuzzy logic of literal classification of very small, small, large, and very large
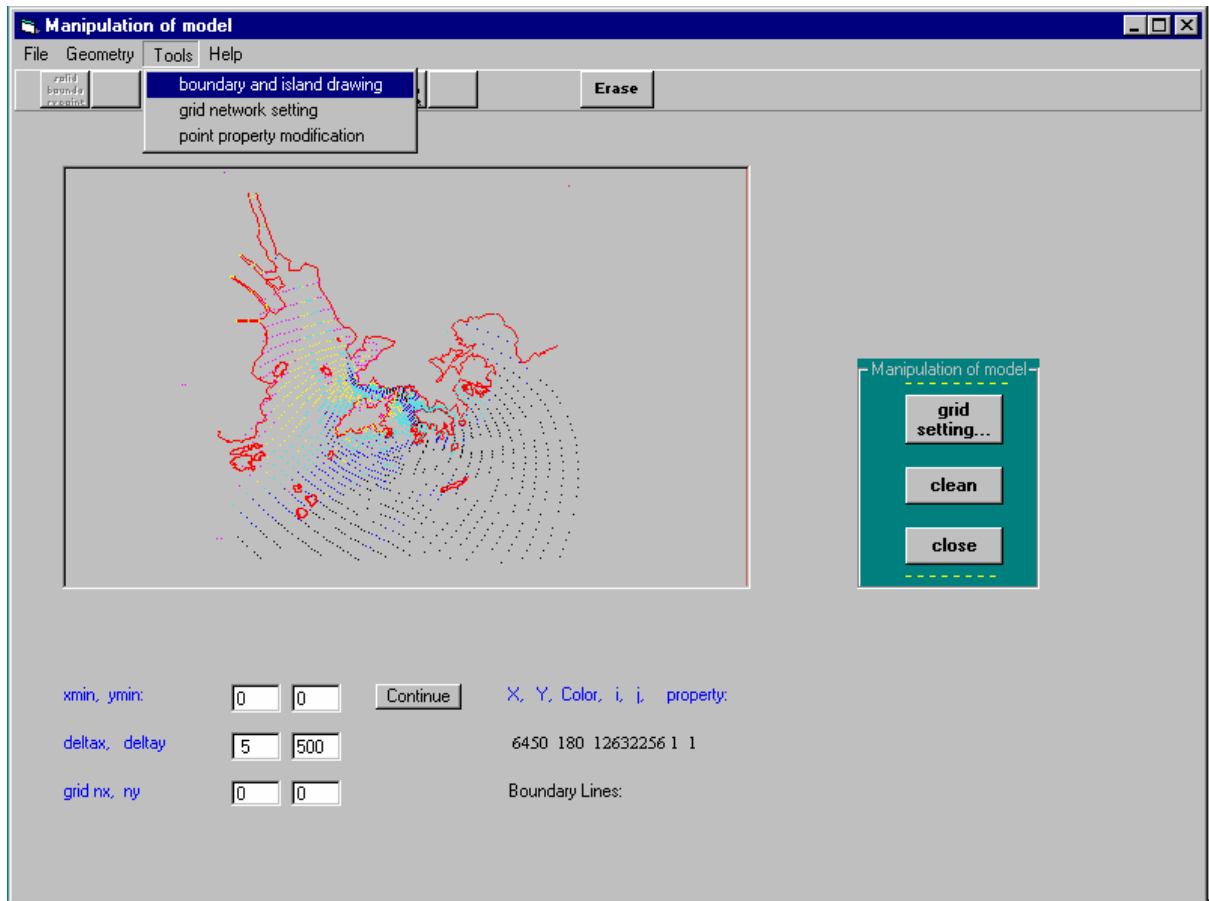
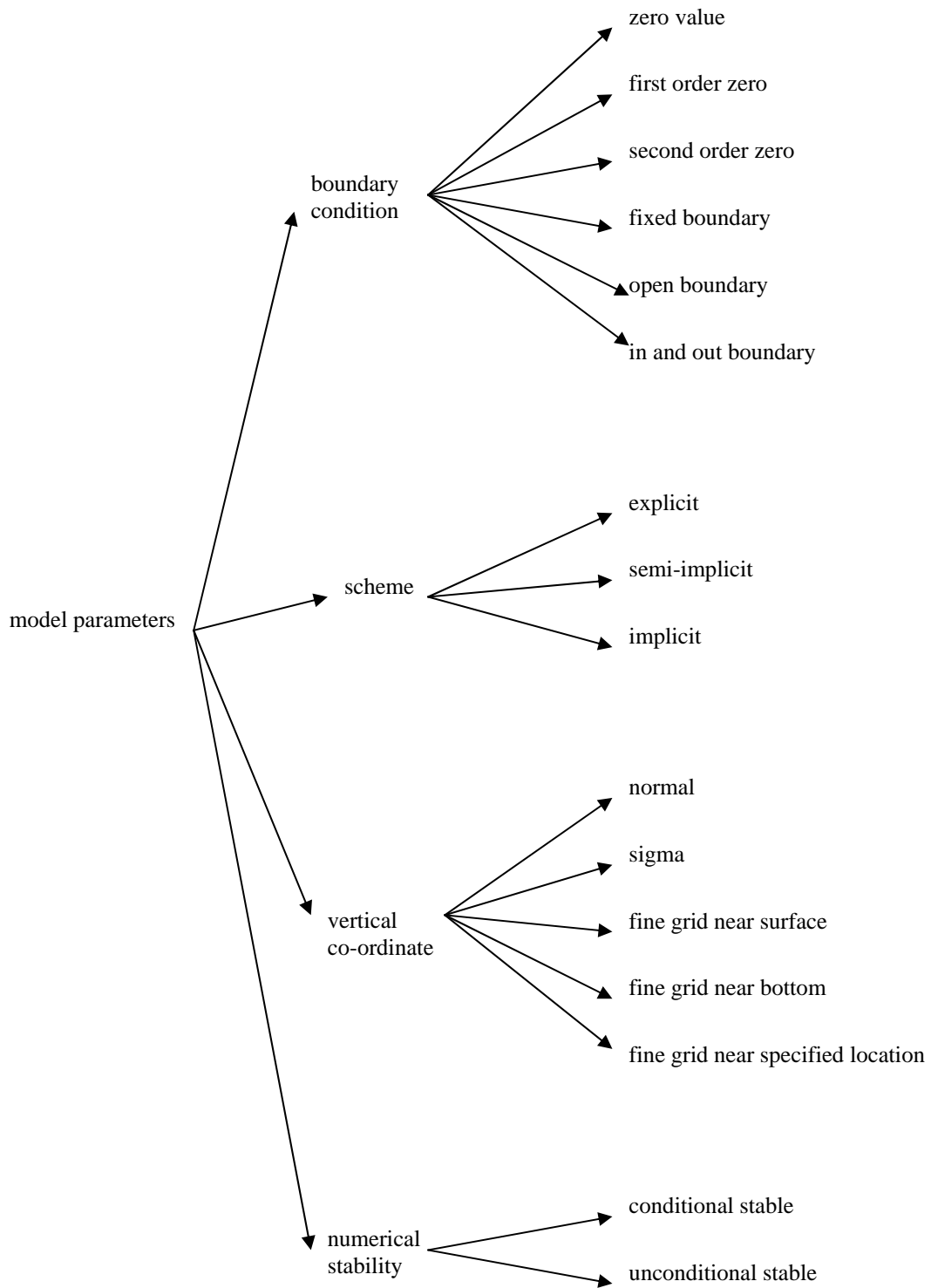Figure 6. A sample screen showing an interactive graphical display of the topography at the Pearl River Estuary

Figure 7. A sample of tree formation in model manipulation