A hybrid reasoning system for supporting estuary modelling

Sara Passone, Vahid Nassehi and Paul W. H. Chung

ABSTRACT

In this paper the development of a Case-Based reasoning system for Estuarine Modelling (CBEM) is presented. The aim of the constructed CBEM system is to facilitate the utilisation of complex modelling software by users who lack detailed knowledge about modelling techniques and require training and assistance to implement sophisticated software effectively. The system is based on modern computing methods and is constructed as a hybrid of three modules which operate conjunctively to guide the user to obtain the best possible simulation for realistic problems. These modules are: a case-based reasoning scheme, a genetic algorithm and a library of numerical estuarine models. Based on the features of a given estuary and the physical phenomenon to be modelled, an appropriate solution algorithm from the system's library is retrieved by the case-based module after a specifically designed reasoning process. The selected model is then analysed and further treated by the genetic algorithm component to find the optimum parameters which can appropriately model the conditions and characteristics of any given estuary. Using these modules the steps that yield the best solution for a problem from the available hydrographic data under a set of specified conditions are explained. This is further elucidated by an illustrative case study which shows the applicability of the present CBEM system under realistic conditions. This case deals with the simulation of salinity distribution in the Tay estuary (Scotland, UK).

Key words | case based reasoning, estuary modelling, genetic algorithm, IT systems

Sara Passone
Department of Chemical Engineering,
Massachusetts Institute of Technology,
Cambridge, MA 02139,
USA

Vahid Nassehi (corresponding author)
Chemical Engineering Department,
Loughborough University,
Loughborough LE11 3TU, UK
Tel: +44 1509 222522
E-mail: *v.nassehi@lboro.ac.uk*

Paul W. H. Chung
Department of Computer Sciences,
Loughborough University,
Loughborough LE11 3TU, UK

INTRODUCTION

Estuaries are under the influence of tides, weather, seasonal river flows and climate. All these aspects control the variation of the water level, salinity, temperature and sediment load and, consequently, the general behaviour of the water body. Furthermore, the human activities that have grown around almost all estuarine areas have had a great impact on the fragile equilibrium of these water systems. The intensive use of estuaries for transportation, food production, waste disposal, flood protection, recreation and other purposes have dramatically increased the environmental stresses affecting these water courses (GESAMP 1990) and modified their morphology (French & Clifford 2000). In addition, the human occupancy of coastal areas is expected to increase worldwide during the next 25 years (Hameedi 1997). In particular, the degradation of estuaries and surrounding areas is expected to accelerate in developing countries where industrialisation is relatively recent and effective legislation for waste disposal and discharge of waste water may not be in place (Kennish 2002).

The physical environment and the behaviour of estuaries have been investigated through a variety of disciplines such as geomorphology, ecology and hydrology over the past decades. However, because of the complexity of such water systems, different approaches have been adopted in these disciplines to investigate the interactions between biological, geological, hydrological, atmospheric and chemical processes within estuaries. Therefore, the focus of investigations carried out by oceanographers, engineers and bio-scientists have been on different aspects of estuarine systems, emphasising a variety of phenomena. As a result, over 40 different definitions (descriptions) of estuaries can be found in the literature, each based on a particular characterisation of estuaries (Perillo 1995). In addition, there are several estuary classifications, each focusing on a unique aspect of the estuarine environment ranging from geomophology and physiography to tidal characteristics, hydrography and sedimentation processes (Dyer 1997). Although an integrated multidisciplinary approach for estuarine science has been recently established, the sources of information and knowledge acquisition in this field are still difficult to access and remain scattered between unrelated disciplines.

Laboratory experiments, field measurements and numerical models are employed mainly by engineers to understand and solve real estuarine problems. In particular, the application of computational hydraulics to estuaries is considered a real breakthrough in attempting to interpret the mechanisms governing estuarine processes. Modern numerical models provide very accurate simulations with minimum cost and in a short time (Thompson 1993). They also enable the investigation and predictive analysis of important aspects of estuarine behaviour which may not be evident from field measurements or analytical evaluations (French & Clifford 2000). Furthermore, their use for investigating the hydrodynamics, sediment movement and mixing processes permits exploration of the effects of a combination of multiple factors. A prime example of this is the use of modelling to develop multiple effluent discharge policies for estuaries (Passone *et al.* 2003).

Although currently used estuarine models involve a large degree of automation it is still necessary for their users to have a high level of expertise in areas of mathematical modelling, physics and estuarine hydrodynamics to utilise them effectively (Dyke 1996). Furthermore the definition and study of estuarine phenomena via numerical modelling is based on an example-by-example approach. Adopting this approach the user needs to have practical experience regarding the appropriate application of numerical models and interpretation of simulation results. To overcome this problem in the last decade a new direction for the development of computational hydraulics has gradually evolved. The purpose behind this has been to extend the accessibility of the numerical models to a broader range of users. The idea is to integrate numerical modelling applications with intelligent reasoning by employing advanced information technology tools (Abbott 1991). In this methodology different artificial intelligence–hybrid systems that facilitate the utilisation and provide guidance at different operational levels of the decision-making process are adopted for the management of water resources. These AI tools handle uncertainty and risk management by using, for instance, the fuzzy set theory (Schulz & Huwe 1997; Mpimpas *et al.* 2001) or genetic algorithm techniques (McKinney & Lin 1994; Aral *et al.* 2001) to define imprecise model parameters. They also focus on the capture and systematic storage of data to process complex information and establish new correlations and patterns in the available model input (Hall 2002). In this respect, AI techniques are used to create computer-aided systems incorporating expertise and existing knowledge to guide and advise the users. Within these knowledge management systems, modelling software is integrated with other information components to work as a single tool for the solution of problems (Cortés *et al.* 2001). Numerical modelling directly benefits from its combination with AI techniques. Characterised by the organisation of the available knowledge and a critical reasoning process, intelligent modelling environments may be created to provide the necessary help for selecting a model that matches any given user goal and the requirements of the problem domain (Knight & Petridis 1992; Sophocleous & Ma 1998; Chau & Chen 2001).

The described approach for the creation of AI–hybrid systems for supporting estuarine modelling has provided the basis for the current research. The system, which is presented in this paper, is implemented using the case-based reasoning (CBR) methodology. Techniques based on "expert systems" have also been used for similar purposes (Chau & Chen 2001). However, the CBR technique provides a more appropriate methodology in dealing with complicated problems arising in estuarine modelling. Obviously practical experience and knowledge of modelling experts cannot be readily converted into an automated system.

Therefore the focus of the present work has been on the codification of simulation results and the assessment of the assumptions which have been used in situations where the interaction of factors determining the estuarine behaviour can be evaluated. It should also be stated that the CBR provides an effective framework for integrating many different AI techniques and paradigms. This feature is essential for retrieving a past case and adapting it to the needs of a new problem. By reflecting the way a modelling expert revises their knowledge when dealing with new circumstances, different types of AI techniques may be combined within the CBR framework to match the characteristics of the problem domain and comply with the principles and assumptions of estuarine modelling theory.

The Case-Based Reasoning for Estuarine Modelling (CBEM) presented consists of three different components: a case-based reasoning scheme, a genetic algorithm and a library of numerical simulation models. These modules, which work as a single tool, are activated to perform specific tasks of the case-based reasoning methodology. With respect to the possible correlation between the features of the estuary and the physical phenomenon to be modelled, the case-based module returns a suitable model from the system's memory. The selected model is then adapted by the genetic algorithm component, which estimates a valid set of model parameters to suit the particular estuarine environment.

The system implements different types of knowledge to drive the model selection. Information related to the practice of estuarine modelling is made available in the description and retrieval components using the rule-based approach and fuzzy set theory.

Simulation of salt intrusion in the Tay estuary is presented to show that the developed approach provides a convenient and effective means for supporting the modelling of estuaries under realistic conditions. The Tay estuary in Scotland, UK, is a meandering mesotidal–macrotidal watercourse with irregular width and depth. Hydrodynamically influenced by its geomorphology, the Tay estuary is classified as "complex" (Buck & Davidson 1997). Therefore the presented case study is complex enough to illustrate various aspects of the system design and the implemented logic.

GENERAL ARCHITECTURE OF THE SYSTEM

As mentioned earlier the developed CBEM consists of three main components (Figure 1). These components (modules) are activated to perform specific tasks of the case-based problem-solving process. The case-based (CB) module allows the user to describe new and past cases (case description). It is also responsible for the retrieval process (case retrieval).

In CBEM a case is divided into two parts: the estuary, which is the object of the investigation, and the related models, each of which is employed to simulate a specific physical phenomenon for that estuary. The estuary description contains indices representing the features of the estuary domain, while the model description includes information about the model characteristics and the estuarine problem simulated. This distinction is due to several practical reasons. The same estuary may have been studied and modelled for different purposes, or a specific estuarine process may have been repeatedly simulated for the same estuary but using different model strategies to satisfy different quality requirements of the results and the specific purposes of the simulations. The separated descriptions for estuaries and models also permit identifying those aspects that contribute to defining a model strategy, which is constructed taking into account the assumptions on the physical and hydrographic behaviour of an estuary, and specific conditions on the problem definition (e.g. cost-effectiveness and accuracy). This distinction also makes the case representation more accessible and clearer to the user. The user then needs to supply the necessary information regarding the component that should be retrieved to speed up the related process (Kolodner 1993).

As the user decides to investigate a problem using CBEM, they enter the feature values of the estuary to be modelled into the estuary description scheme. They then define the type of problem and the purpose of the investigation. The retrieval process begins with the search engine selecting from the system's case base only those cases for which the current problem has been previously modelled. For each of these past cases, the retrieval mechanism then computes two similarity ratings based on their estuary and model descriptions, respectively. Eventually, the user is presented with a list of past cases graded with respect to the estuary description as well as the

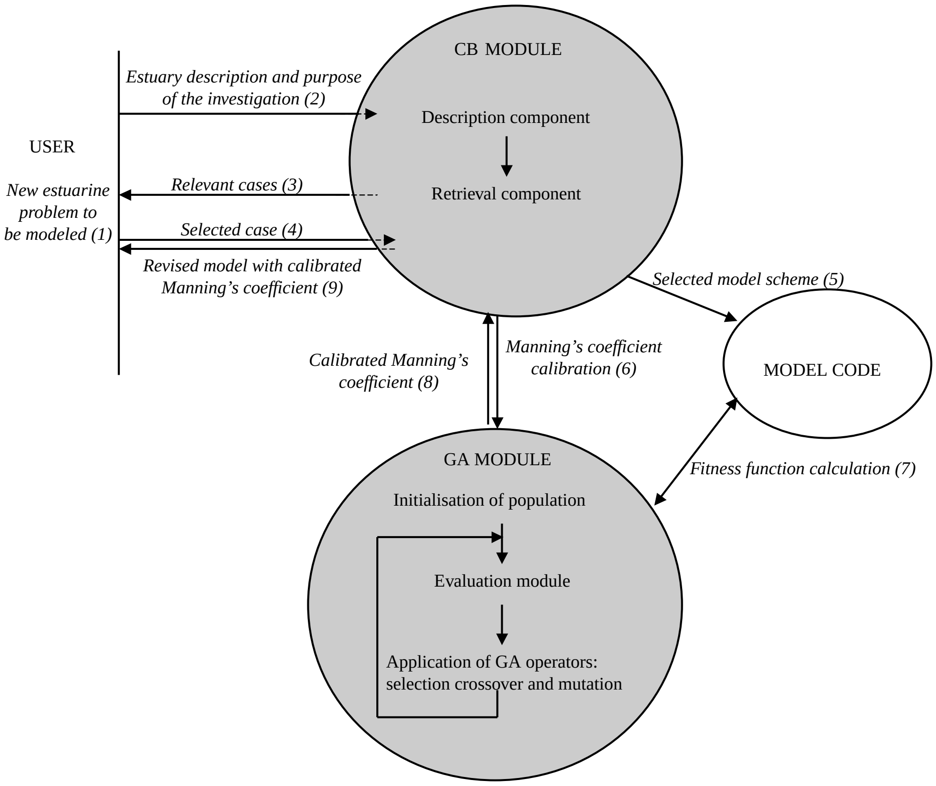

Figure 1 | CBEM architecture.

problem definition. The user is responsible for the final selection based on personal judgment.

Once the model scheme is selected, it is given to the genetic algorithm module (GA) which is responsible for the optimisation of the model parameters. This module is developed by combining the classical evolutionary approach with problem-specific information. Modifying versions of the classical genetic operations of initialisation, selection, crossover and mutation have been designed to incorporate knowledge related to estuaries and the calibration of estuarine models. Furthermore, the present scheme benefits from cooperation with the CB module by including in the initial population parameter values from the most similar cases. The use of knowledge augmented operators (Goldberg 1989) and case-based initialisation (Grefenstette 1987) improves the search performance, addressing the search towards those zones of the search space that more likely contain the suitable solutions. The CBEM procedure terminates with the return to the user of the model scheme, retrieved from among the past cases as the best match, equipped with a new set of parameters to guarantee a satisfactory performance.

CB MODULE – DESCRIPTION COMPONENT

Estuary description

In order to identify a model within the CBEM's case base that is suitable to simulate the salinity distribution problem in the Tay estuary, the user needs to describe the estuary through a case description form (Figure 2). In CBEM the organisation and representation of estuaries are assessed on the basis of the classification proposed by Dyer (1997), i.e. water circulation, mixing process and stratification. By combining the information contained in these schemes, the estuary description is here organised in terms of physical and hydrographic features. The physical features represent the dimensions of an estuary. They are the geomorphological type, the tidal range, the estuarine total area, the intertidal area, the maximum, minimum and average widths, the average

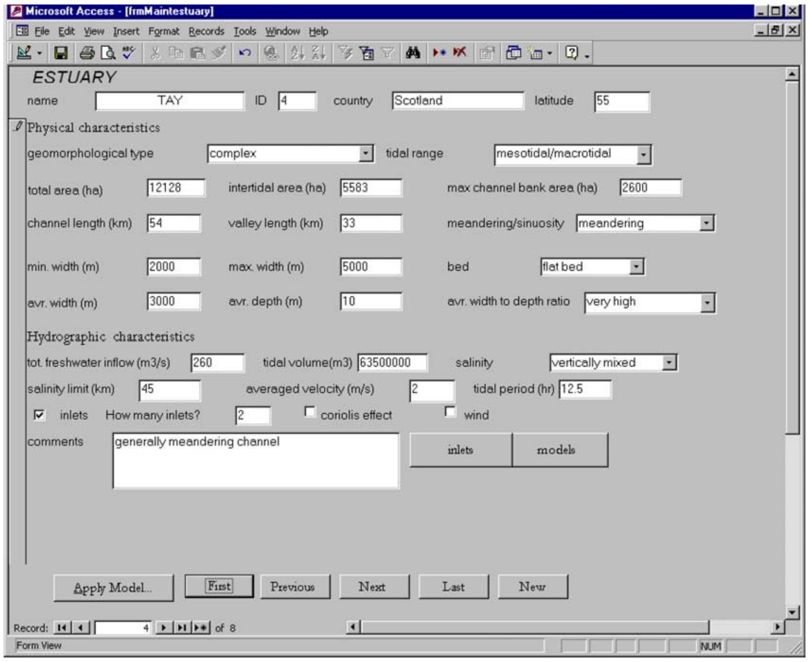

Figure 2 | Description of the Tay estuary, Scotland, included in CBEM.

depth, the channel length, the valley length, the grade of estuary meandering and the bed shape. The hydrographic characteristics are the freshwater flow, the tidal flow, the salinity, the limit of the salinity intrusion within the estuary channel and the average longitudinal velocity. There are also included indices for the wind, the Coriolis forces and the number of estuarine inlets. However, the information contained in these classifications needs to be carefully combined. This is because these classifications focus on specific aspects of the estuarine environment and do not take into account the possible interdependence between different aspects of an estuary. Therefore, this estuary description, which is valid for both past and new cases, is designed to organise systematically the information scattered through these classifications into formal and meaningful indices, suitable for the case-based reasoning process. Some physical estuary's features are defined using qualitative symbols in order to facilitate the indexing and retrieval processes. By using fuzzy set theory, the width to depth ratio and the degree of an estuary's sinuosity are expressed according to a qualitative scale. In addition, some indices such as the geomorphological type, the tidal range and the salinity stratification, as object symbols (Chung & Inder 1992) are defined according to the description provided by the estuaries' classification schemes.

While the estuary description scheme is employed for representing past as well as new cases, the model description is only a part of the cases already within the system's memory. The proposed scheme for model description (Figure 3) was designed based on "Classification of the models of tidal waters" by Hinwood & Wallis (1975) and "Guidelines of the use of computational models in coastal and estuarial studies" by Lawson & Gunn (1996). These authors have used a large number of instantiated models to classify different estuaries. The model is described in terms of: the type of problem, dimension, numerical technique, model assumptions (e.g. presence of wind and Coriolis force, etc.), dispersion and Manning's coefficients. The values of these two coefficients are included as they may be used during the adaptation phase. They may be included into the initial population of the genetic algorithm routine, if the corresponding model is retrieved by the system as being the most appropriate to the new problem. It should be noted that the criteria described for the model classification and selection are not exhaustive and the procedure can be further developed by including other factors such as the nature of the turbulence model used in

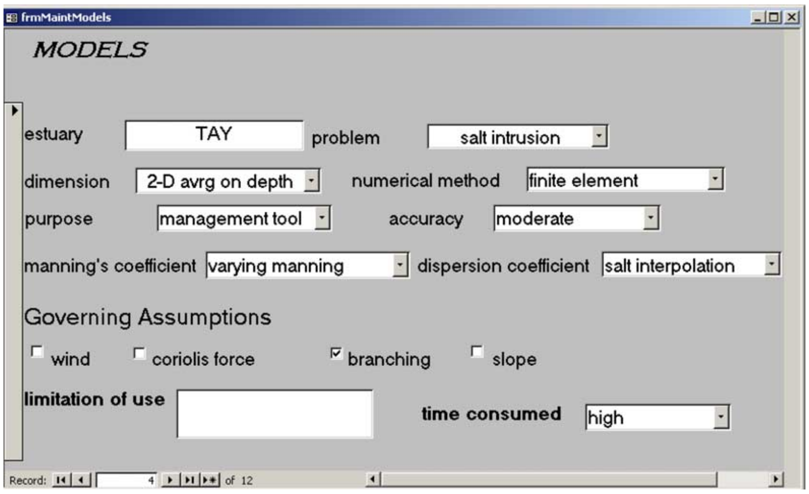

Figure 3 | Example of model represented with respect to the model description scheme provided by CBEM.

the governing model equations, alternative friction slope to the Manning's coefficient and type of scheme utilised for the numerical simulations.

Considerations related to the model strategy are also part of the model description. These features are accuracy, times required and simulation purpose. During the retrieval process these attributes are essential to estimate the appropriateness of a model when the user's requirements for efficiency and accuracy are taken into account. For instance, a model may provide a sufficiently correct simulation procedure but may be inappropriate according to the aim of the investigation as far as accuracy, time required and simulation purpose are concerned.

CB MODULE – RETRIEVAL COMPONENT

After entering the new case according to the estuary description scheme, the user is required to specify in the model selection screen (Figure 4) the type of problem to be simulated and the purpose of the undergoing investigation. These attributes, whose values are identical to those used for the homonymous indices in the model description, provide the necessary information to calculate the suitability of models previously used. For instance, in the case of the Tay estuary the user specifies "salinity distribution" as the type of problem and the purpose of the investigation as checking the suitability of the model as a "management tool".

Once the indices of the model selection screen have been chosen, the retrieval process can be activated. The degree of similarity is computed using the nearest-neighbour matching procedure, which evaluates the similarity in two stages. The first similarity rating between a new and a past case is defined based on the values of the physical and hydrographic characteristics contained in their estuary descriptions. A second score is then calculated with respect to the type of investigation to be conducted, the accuracy and simulation time required. The similarity is measured using a fuzzy approach if the attributes are described on a quantitative scale (i.e. ratio of the total area to the intertidal area). Alternatively, if the descriptors are expressed

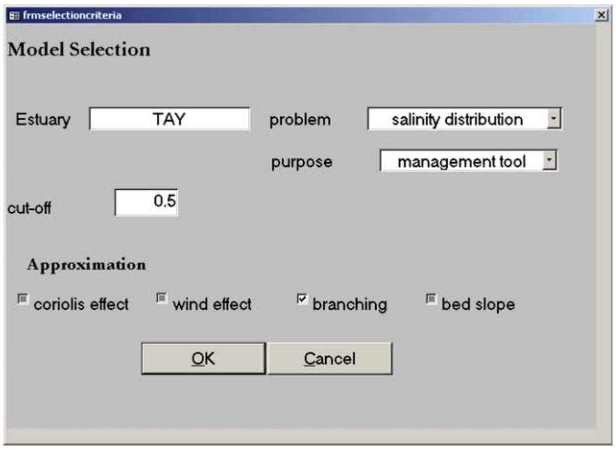

Figure 4 | Model selection screen employed by the user to specify the new problem case.

qualitatively (i.e. the degree of meandering or the model purpose), the matching criterion consists of computing the distance between the two symbols.

The heuristic criteria of exclusion and preference expressed with respect to the model dimension are also employed during the retrieval process. The model dimension, which depends on the type of problem as well as the estuary's physics, needs to be chosen so that the physical phenomenon under investigation is well represented without underestimating or over-sizing the problem domain. Cases are not retrieved from the case base if the dimension of the related models is considered inappropriate for the type of problem in hand (criterion of exclusion). Some cases are also preferred over others if their model dimension as evaluated by the system is more suited to simulate the current problem (criterion of preference).

The retrieval procedure implemented in the present scheme uses different sets of matching and importance criteria according to the type of estuarine phenomenon to be modelled. Currently different case ranking procedures are implemented in CBEM for the problems of salinity distribution and salt intrusion.

SIMILAR ESTUARIES

In the retrieval process the similarity based on the values of the features of the estuary description is computed first. The similarity of the new case's estuary is estimated as follows:

1. Select a set of 7 features F from the estuary description: the ratio of the average width to the average depth (a), the geomorphological estuary type (b), the tidal range (c), the meandering rate (d), the ratio of the total area to the intertidal area (e), the ratio of maximum bank channel area to the intertidal area (f) and the ratio of the channel length to the average depth (g).
2. Assign the degree of relevance $W=1$ to (a), 0.75 to (b) and (c) and 0.5 to (d), (e), (f) and (g).
3. Determine the similarity values $S = \mathrm{sim}\left(F_k^I, F_k^R\right)$, with sim as the similarity function and I and R referring to the input and retrieved cases, respectively.
4. Normalise the match aggregate score $= \frac{\sum_{k=1}^{7} w_k s_k}{\sum_{k+1}^{7} w_k}$.

Figure 5 shows the case of the Tay estuary and the assessment of its similarity with the British estuaries of

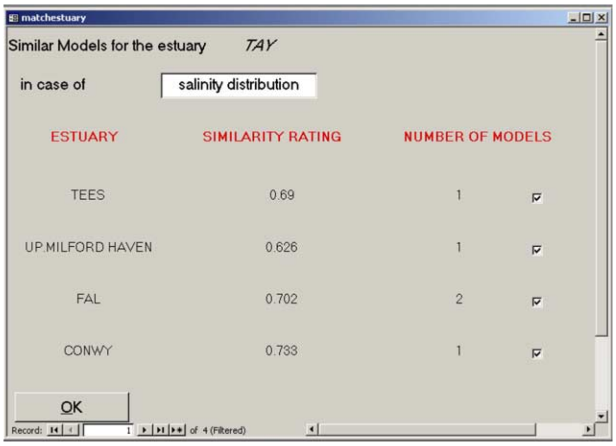

Figure 5 | Values of Tay estuary's similarity with various British estuaries.

Tees, Upper Milford Haven, Fal and Conwy. For each retrieved estuary CBEM gives the value of the similarity rating and the number of models previously employed to simulate the salinity distribution in each estuary.

At this point the user must select those cases that he/she would like to see through the second phase of the retrieval process. For the purpose of the present discussion, all four retrieved estuaries are admitted. The second half of the process consists in the computation of the second similarity score that quantifies the appropriateness of each retrieved model with respect to the specified purpose of the investigation (i.e. "management tool").

Similar models

The retrieval process continues with the computation of the adequacy of each retrieved model. This is evaluated through a pre-determined set of match values that rank the accuracy, simulation time consumed and purpose of the model based on the investigation aim of the new problem (i.e. management purpose). The procedure for calculating the likelihood of each model to fulfil the user's requirements is described below.

1. Exclude case models with respect to the model dimension.
2. Select the following features from the model description: purpose (h), accuracy (i), simulation time (j).
3. Apply a set of pre-determined rules to establish the functional role M_k of each feature with respect to the purpose of the current investigation.

4. Assign a grade of relevance $P = 0.75$ to (h) and 0.5 to (i) and (j).
5. Apply the criterion of preference with respect to the model dimension. If a model is "preferred", assign the value 1 to the match value M_k and to the grade of relevance P.
6. Normalise the match aggregate score $= \frac{\sum_{k=1}^{N} P_k M_k}{\sum_{k=1}^{N} P_k}$ with N equal to 4 if the criterion of preference is valid; otherwise N is equal to 3.

At the end of the retrieval process the selected cases are classified with respect to both the first and second similarity scores (Figure 6). However, only 4 out of the initially retrieved 5 cases are in this list. The case related to the Tees estuary is withdrawn from the set. This case is eliminated because of the principle of exclusion, which tells us: "If the estuary has inlets and at least one of the inlets is a 'branch' then eliminate all 1-D models". It should be noted that, by using appropriate mass and momentum balance relationships at the branch junctions, 1D models can still be applied to branching estuaries (Nassehi & Williams 1986). Therefore the principle of exclusion on the basis of branching in the flow channel only applies to the cases where the model has not been generalised by the inclusion of appropriate equations for the junctions.

Although the Fal estuary is not considered by CBEM to be similar to the Tay estuary, in terms of its physical and hydrological characteristics (Figure 6), the use of a 1D network model which involves "low" simulation time, "moderate" accuracy and provides data regarding "water quality" (Figure 6) for investigating salinity distribution in the Tay is the logical choice.

GA MODULE

In numerical models, case adaptation is essentially based on adjusting parameters to ensure that the model accurately simulates the real behaviour of the new estuary. Calibration of numerical models for estuaries consists of determining the values of these parameters that provide the better agreement between the model simulation and the observed hydrodynamics of these water systems. The selection of model parameter values is affected by physical phenomena characterising the specific water system considered as well as scale effects due to approximations introduced during the model development. Because their best values cannot be obtained by direct field measurements, special techniques, based on the minimisation of the difference between simulated values and the observed data, need to be employed.

In the present CBR system a genetic algorithm is used to identify appropriate values for Manning's coefficient. This parameter is utilised to represent the bed resistance to the flow of water in the hydraulic equation of motion. This coefficient reflects the variations of the physical and geometrical characteristics of the watercourse. Manning's coefficient in estuaries typically varies within a range between $0.011-0.060\,\mathrm{m}^{1/3}\,\mathrm{s}$ (Chow 1973). For numerical models where the problem domain is discretised into elements (up to several hundreds), the resistance to the

retrievedmodels

ESTUARY	DIMENSION	ACCURACY	TIME	PURPOSE	ESTUARY SIM.	MODEL SIM.
UP.MILFORD HAVEN	2-D moving	high	high	research	0.626	0.409
FAL	1-D network	moderate	low	water quality	0.702	0.545
FAL	2-D avrg on depth	high	high	research	0.702	0.409
CONWY	2-D avrg on depth	high	moderate	water quality	0.733	0.491

Record: 1 of 4 (Filtered)

Figure 6 | Cases retrieved by CBEM classified with respect to their model characteristics and the rates of similarity with the Tay estuary and the purpose of the new investigation.

flow is expressed by associating with each section a specific value of Manning's coefficient. Thus, calibrating Manning's coefficient in a numerical hydrodynamic model of an estuary means to find the set of Manning's coefficient values that gives realistic simulations. It must be noted that, because of the interdependency of Manning's coefficient values on each other, just changing a Manning's coefficient value for one section of the domain may result in the alteration of the entire model performance and the quality of the output.

Model calibrations have been traditionally carried out either manually or using numerical optimisation programs. However, both manual and computer-based parameter optimisation require an experienced modeller (McDowell & O'Connor 1977). Furthermore, some practical parameter spaces, such as the domain of possible sets of Manning's coefficients in hydrodynamic models, are too large to be investigated either manually or even using computer-based numerical algorithms (Goldberg 1989).

The GA module carries out the calibration of Manning's coefficient independently by combining the classical evolutionary scheme with problem-specific information in the form of heuristic rules and case-based reasoning principles. The classical genetic operations of initialisation, selection, crossover and mutation are modified to incorporate practical information about the estuarine model calibration. This implementation narrows down the areas of the search space where the best set of parameters is more probably included. A considerable reduction of the necessary computational time is then obtained.

Initialisation

The chromosomes are represented using the decimal base. As Manning's coefficients differ from one value to another only in the last two digits, the chromosomes are expressed as integers corresponding to the second and third decimal places of Manning's numbers. This representation is more practical than the classical binary code since, to preserve the accuracy of modelling, the number of elements in a discretised domain is usually high (up to several hundreds). Therefore, with a large number of Manning's coefficients the use of integers for the genes significantly facilitates the passage to and from the phenotypical representation and the transformation by genetic operators.

Furthermore, the initialisation procedure is not based on the classical method of randomly generating chromosomes. This practice is not considered a feasible choice for the domain of estuarine modelling as Manning's coefficient remains the same, or varies very little, for adjacent elements in a discretised domain. The reason for this is that the resistance to the flow changes with respect to the variation of the estuarine physical characteristics and reaches with similar physical features are expected to have similar values for Manning's coefficient. Therefore, adjacent elements are likely to have similar values for Manning's coefficient. By using randomly generated coefficients there is the danger of obtaining unrealistic simulations. Instead, the present GA scheme considers each chromosome divided into a number of segments corresponding to the zones of the estuary with specific physical characteristics. Based on this chromosome's structure, a value of Manning's coefficient is randomly generated for each segment and assigned to the genes of the corresponding segment (zonation option) (Figure 7).

The observation that the flow resistance generally decreases towards the estuary mouth is also taken into account during the initialisation. Based on this evidence, the GA program sorts the alleles (i.e. the variations in the genes regarding their values) of chromosomes in descending order, with lower values for genes that correspond to elements of the domain allocated towards the estuary mouth (scaling option). In the example provided in Figure 7, the values of the genes gradually decrease from zone A towards zone B, which include the estuary's head and mouth, respectively.

The last feature implemented for generating the initial population consists of seeding the cluster with appropriate

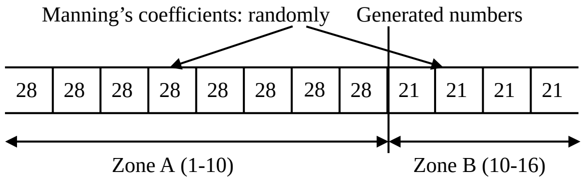

Figure 7 | Example of chromosome generated using the decimal base and initialised according to the zonation and scaling option.

Manning's coefficient series selected from the system's case base (case information). Based on the principle that similar problems should have similar solutions (Louis & Johnson 1997), estuaries that do not significantly differ from one another should have similar sets of Manning's coefficients. The sets are, however, adapted to suit the discretisation scheme employed for the estuary under investigation. The use of case information is limited to 10% to avoid premature convergence and ensure the necessary population diversity.

GA operators

The GA operators of selection, crossover and mutation are also designed to incorporate concepts from the theory of estuarine calibration for the purpose of Manning's coefficient optimisation.

Starting from the initial population the subsequent generations are formed by selecting the chromosomes according to their fitness. The fitness of the chromosomes can be computed by estimating the discrepancy (ρ) between the water surface elevations (H_m) measured at different locations within the estuary, and their corresponding simulated values (H_s). Each sampling station j is characterised by a set of experimental data corresponding to the water surface elevations observed at different time levels, indicated by n. Denoting the total number of sampling stations by J and the total number of samples, collected at each station during a tidal period, by N, the series of all measured water surface elevations can be represented as $H_m = \left\{(h_m)_n^j, j = 1, \ldots, J; n = 1, \ldots, N\right\}$ and the set of all simulated values as $H_s = \left\{(h_s)_n^j, j = 1, \ldots, J; n = 1, \ldots, N\right\}$.

Hence, the discrepancy between H_m and H_s is given as

$$\rho(H_s, H_m) = \left[\sum_{j=1}^{J}\sum_{n=1}^{N}\left((h_s)_j^n - (h_m)_j^n\right)^2\right]^{1/2}. \quad (1)$$

The fitness of each chromosome is calculated as the reciprocal of ρ:

$$\varphi_I = \frac{1}{\rho}. \quad (2)$$

In order to find which chromosome gives a maximum for Equation (2), water surface elevations for all chromosomes in each generation must be simulated. Hence, h_s at each station j for the time levels n is calculated.

The selection procedure implemented consists of keeping 10% of the best chromosomes (i.e. with the highest fitness values) in the next generation (i.e. elitist approach) and having the other 90% of the next generation randomly reproduced according to their fitness values (i.e. roulette wheel) and then transformed by crossover and mutation in order to introduce diversity into the population. This stops the search converging too quickly and more of the search space is explored.

The present scheme also contains different forms of the more common random mutation and crossover. The crossover and mutation operators are devised to guide the search towards chromosomes with a real physical meaning for estuarine calibration. Therefore, the traditional genetic operators are modified according to the previously made observations of adjacent genes representing adjacent elements and chromosome segments corresponding to specific estuary zones. The crossover operator swaps between chromosomes segments which correspond to specific estuary zones. The number of cut points in a chromosome is randomly chosen each time the crossover operator is applied (Figure 8). The mutation operator implemented here is based on the concept that close elements are generally characterised by similar Manning's coefficients.

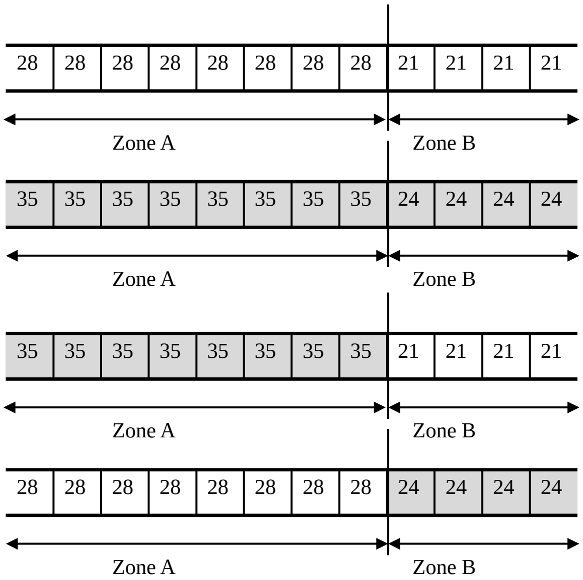

Figure 8 | Crossover operator applied between chromosome segments corresponding to specific estuary zones.

Thus, the chromosomes are mutated by changing the value of a randomly chosen gene and its closest neighbours (Figure 9).

The automated GA calibration described here works on the basis of matching the observed and simulated water surface elevations and it may lead to a "force fitting" of the results. This is because in complex water systems such as estuaries discrepancies between the simulated and observed data can arise from a variety of diverse sources including uncertainties in input data, model structure or model parameters, to name a few. Therefore there is a danger that "force fitting" of simulated results based on altering a single factor of bed roughness may artificially mask the inefficiencies of the model structure or uncertainties of the input data. To avoid this, after the determination of the "best" set of roughness coefficients at the end of the calibration stage the true physical significance of the obtained values should be verified. At this stage the values of the roughness coefficient are kept constant and a situation different to the calibration event is simulated. The comparison between this result with the observed data provides a measure of the "safety" of the roughness coefficients determined through the calibration.

GA calibration for the Tay estuary

Once the 1D network model has been selected for application in simulating the salinity distribution problem in the Tay estuary (Figure 6), it is given to the adaptation component to be adjusted by the GA module. The GA module computes the set of Manning's friction coefficients that are appropriate for the particular hydrodynamics and geomorphological characteristics of the Tay estuary. The present calculations are for a typical spring tide (12 June 1972), using the measured water surface elevations at

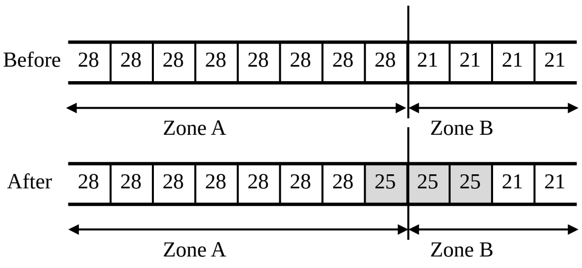

Figure 9 | Mutation operator applied to a gene and its closest neighbours.

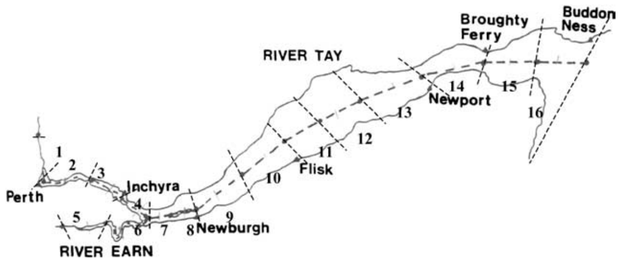

Figure 10 | Discretisation into 16 elements of the Tay estuary domain.

Buddon Ness (Figure 10) and the fresh water inflow at the estuary head as boundary conditions. The estuary domain is discretised into 16 elements within which the equations of motion and continuity are solved to obtain water surface elevations at the nodal points of the elements (Figure 10). The detailed derivation of the mathematical model and the finite element solution scheme have been presented elsewhere (Bikangaga 1993) and are not repeated here.

The genetic algorithm calibration of the model is executed with a population of 30 individuals and the rate of crossover and mutation equal to 0.5 and 0.01, respectively. The genetic algorithm is run for 15 generations. The estuary is divided into two zones. The first zone corresponds to the area of the estuary between elements 1 and 10, while the second zone corresponds to that part of the channel from element 11 towards the estuary mouth. Based on this partition of the estuary the chromosome population is initialised using the zonation and scaling options. These chromosomes are then transformed by the modified mutation and crossover operators. Only one set of Manning's coefficients from the case library is included in the initial population, which is the set of parameters employed for the simulation of salinity distribution in the Fal estuary using the 1D model network.

The calculation of the fitness function is based on the minimisation of the discrepancy between the model output and the observed data at Inchyra and Newburgh (Figure 10). These locations are selected because the propagation of a tidal wave towards the estuary head is accompanied by significant deformation of its shape at these two stations. In Figure 11 the simulated water surface elevations, generated

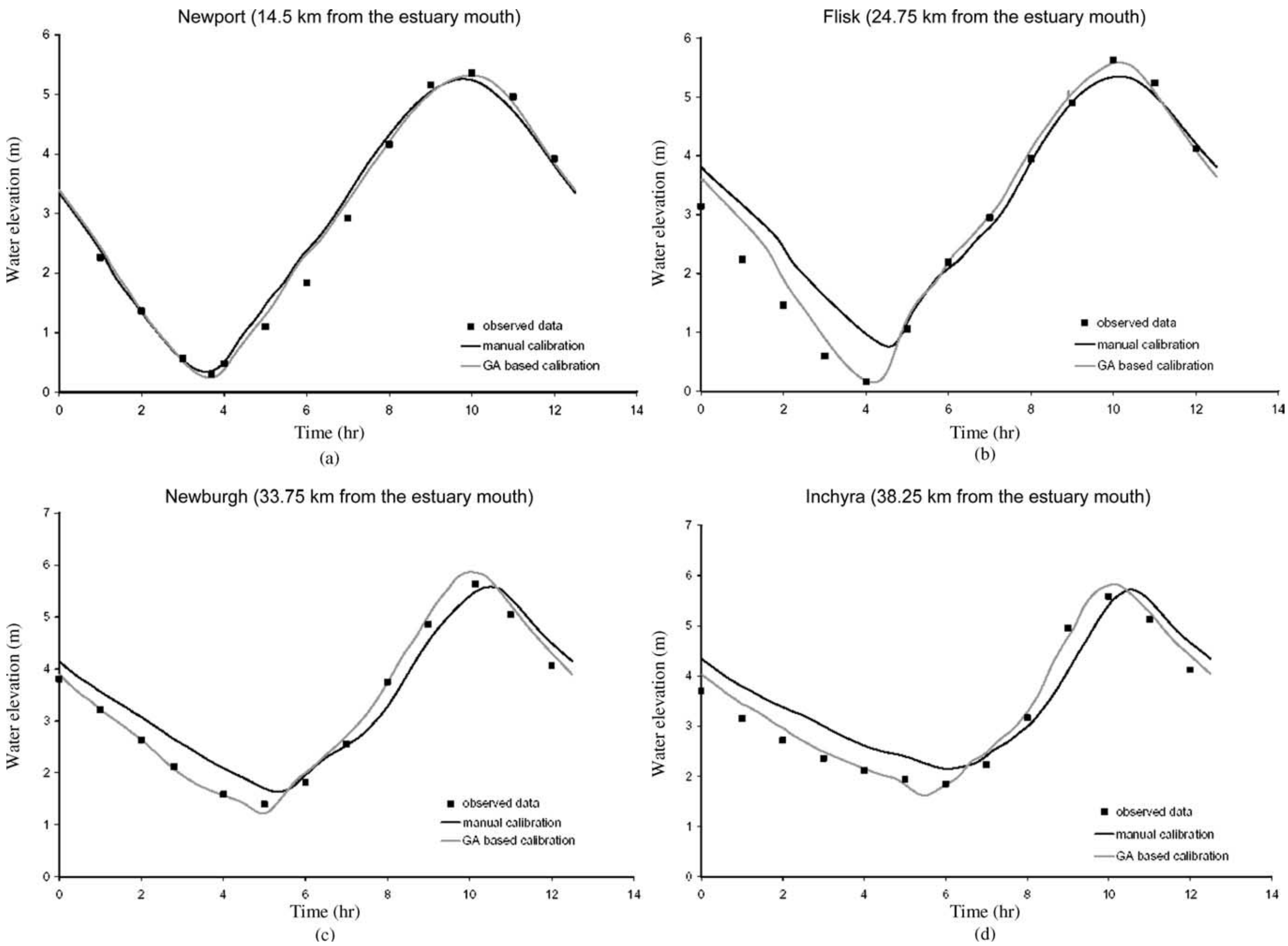

Figure 11 | Water surface elevations simulated using the genetic algorithm optimisation scheme and the manual calibration at four locations on the Tay estuary: (a) Newport, (b) Flisk, (c) Newburgh and (e) Inchyra (spring tide of 12 June 1972).

using the set of Manning's coefficients selected by the GA module, are presented for the stations of Newport, Flisk, Newburgh and Inchyra. In addition, the observed and simulated water surface elevations, obtained by manual optimisation of the model at the described locations, are also shown in Figure 11.

Both manual and GA simulations confirm that the tidal wave propagates with a progressive deformation of its shape towards the estuary's head (i.e. Inchyra). However, comparison between the observed data and the simulated water surface elevations for the manual and the GA-based calibrations shows that the model optimised by the GA routine uses an automatic, and hence a more convenient, procedure. In particular, while the two sets generate similar water elevations at Newport, the GA-based calibration copes more efficiently with the deformation of the tidal wave, which takes place between the extensive floodplain and the estuary head. There are only small discrepancies between the observed and simulated water surface elevations, obtained using the GA-based calibration at Flisk, Newburgh and Inchyra.

As mentioned earlier it is necessary to verify the validity and physical significance of the calibrated roughness coefficients in a further verification step. In the present work this is done by comparison of the model results with the observed values of the water surface elevation in the Tay estuary for the neap tide of 20 June 1972. The results of this comparison are shown in Figure 12 (a–d).

The superiority of the GA-based calibration over the trial and error optimisation is also demonstrated by

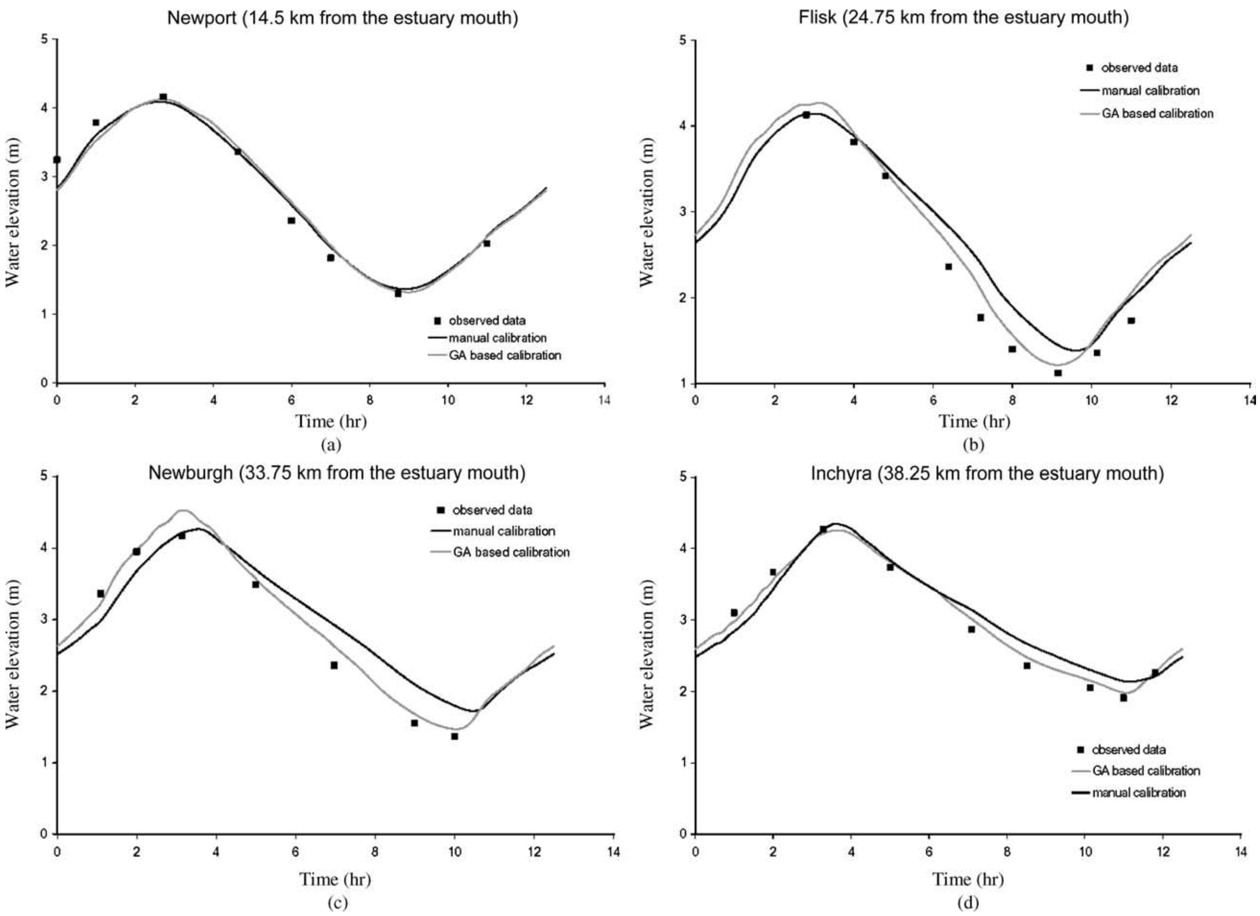

Figure 12 | Water surface elevations simulated using the genetic algorithm optimisation scheme and the manual calibration at four locations on the Tay estuary: (a) Newport, (b) Flisk, (c) Newburgh and (e) Inchyra (neap tide of 20 June 1972).

considering the time necessary to carry out these two processes. The manual calibration of a one-dimensional hydrodynamic model based on the division of the tidal flow channel into 100 segments (elements) requires days of work while the GA-based calibration takes 10 h of CPU time in a shared Sun workstation.

CONCLUSIONS

The described "hydroinformatics" system is shown to have the potential of providing a very effective tool in assisting users to select, apply and obtain predictive data about realistic estuarine problems. Although the system assumes that the user has only a minimal knowledge concerning mathematical modelling it is capable of generating detailed output which can easily lead to reasoned analysis and ultimately the appropriate management decisions. It is also shown that using module-based algorithms lengthy operations such as the optimisation of physical parameters can be achieved in a very short time by utilising effective techniques such as genetic codes. Application of the developed system to a real case has given an example which elucidates the workings of the present CBEM.

At its present state of development the described CBEM relies on a relatively limited number of previously analysed cases. However, it can be readily extended to include new instantiated cases as they become available. Obviously such an extension will enhance the applicability and usefulness of the presented case-based reasoning system.

ACKNOWLEDGEMENT

The authors wish to express their gratitude to Mr Navraj S. Hanspal for reading the manuscript and making valuable comments about its contents.

REFERENCES

Abbott, M. B. 1991 *Hydroinformatics – Information Technology and the Aquatic Environment*. Avebury Technical, Aldershot, UK.

Aral, M. M., Guan, J. B. & Maslia, M. L. 2001 Identification of contaminant source location and release history in aquifers. *J. Hydrol. Engng.* **6** (3), 225–234.

Bikangaga, J. H. 1993 *Mathematical Modelling of the Hydrodynamics and Transport of Soluble Reactive Pollutants in Narrow Tidal Rivers.* PhD Thesis, Loughborough University Of Technology, UK.

Buck, A. L. & Davidson, A. C. 1997 *An Inventory of UK Estuaries*. Joint Natural Conservation Committee, Peterborough.

Chau, K. W. & Chen, W. 2001 A fifth generation numerical modelling system in coastal zone. *Appl. Math. Modell.* **25**, 887–900.

Chow, V. T. 1973 *Open-Channel Hydraulics*. McGraw-Hill, New York.

Chung, P. W. H. & Inder, R. 1992 Handling uncertainty in accessing petroleum exploration data. *Rev. Institut Francais du Petrole* **47** (3), 305–314.

Cortés, U., Sànchez-Marrè, M., Sangüesa, R., Comas, J., Roda, J. R., Poch, M. & Riaño, D. 2001 Knowledge management in environmental decision support system. *AI Commun.* **14**, 3–12.

Dyer, K. R. 1997 *Estuaries A Physical Introduction*. John Wiley & Sons, New York.

Dyke, P. 1996 *Modelling Marine Processes*. Prentice-Hall, London.

French, J. R. & Clifford, N. J. 2000 Hydrodynamic modelling as a basis for explaining estuarine environmental dynamics: some computational and methological issues. *Hydrol. Process* **14**, 2089–2108.

GESAMP 1990 *The State of the Marine Environment*. United Nations Environment Programme, Nairobi.

Goldberg, D. E. 1989 *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA.

Grefenstette, J. J. 1987 Incorporating problem specific knowledge into genetic algorithms. In *Genetic Algorithms and Simulated Annealing* (ed. L. D. Davis), Pitman, London. pp. 42–61.

Hall, J. W. 2002 Random set uncertainty handling in hydroinformatics. In *Proc. 5th International Conference on Hydroinformatics* (eds I. D. Cluckie, D. Han, J. P. Davis & S. Heslop), IWA Publishing, London, pp. 1388–1393.

Hameedi, M. J. 1997 Strategy for monitoring the environment in the coastal zone. In *Coastal Zone Management Imperative for Maritime Developing Nations* (ed. B. U. Haq), pp. 111–142. Kluwer, Amsterdam.

Hinwood, J. B. & Wallis, I. G. 1975 Classification of models of tidal waters. *J. Hydraulics Div., Proc. ASCE* **101** (HY10), 1315–1351.

Kennish, M. J. 2002 Environmental threats and environmental future of estuaries. *Environ. Conserv.* **29** (1), 78–107.

Knight, B. & Petridis, M. 1992 *FLOWES*- an intelligent computational fluid-dynamics system. **5**, pp. 51–58. Engineering Applications of Artificial Intelligence, Pergamon.

Kolodner, J. 1993 *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, CA.

Lawson, J. & Gunn, I. J. 1996 *Guidelines for the Use of Computational Model in Coastal and Estuarial Studies*. *Report SR 450*. HR Wallingford, Oxon, UK.

Louis, S. J. & Johnson, J. 1997 Solving Similar Problems Using Genetic Algorithms and Case-Base Memory. In *Proceedings of the Seventh International Conference on Genetic Algorithms*, Morgan Kauffman. San Mateo, California, pp. 101–127.

McDowell, D. M. & O'Connor, B. A. 1977 *Hydraulic Behaviour of Estuaries*. Macmillan, London.

McKinney, D. C. & Lin, M. D. 1994 Genetic algorithm solution of groundwater-management models. *Wat. Res. Res.* **30** (6), 1897–1906.

Mpimpas, H., Agnastopoulos, P. & Ganoulis, J. 2001 Modelling of water pollution in the thermaikos gulf with fuzzy parameters. *Ecol. Modell.* **142** (1–2), 91–104.

Nassehi, V. & Williams, D. J. A. 1986 Mathematical model of Upper Milford Haven – a branching estuary. *Estuarine, Coastal Shelf Sci.* **23**, 403–418.

Passone, S., Das, D. B., Nassehi, V. & Bikangaga, J. H. 2003 Design of discharge policies for multiple effluent sources and returning pollutant scenarios in a branching estuary. *Estuarine, Coastal Shelf Sci.* **56**, 227–237.

Perillo, G. M. E. 1995 *Geomorphology and Sedimentology of Estuaries*. Elsevier, Amsterdam.

Schulz, K. & Huwe, B. 1997 Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach. *J. Hydrol.* **201** (1–4), 211–229.

Sophocleous, M. & Ma, T. S. 1998 A decision support model to assess vulnerability to salt water intrusion in the Great Bend Prairie Aquifer of Kansas. *Ground Water* **36** (3), 476–483.

Thompson, G. 1993 Mathematical models and engineering design. *J. Inst. Water Environ. Mngmnt.* **7**, 18–23.