The Use of a Genetic Algorithm in the calibration of estuary models

S. Passone[1] and **P.W.H. Chung**[2] and **V. Nassehi**[1]

Abstract. This paper describes an artificial intelligence (AI) system for estuarine model design. It is created by the combination of case-based reasoning and genetic algorithm techniques. This application aims to make the utilisation of complicated and expensive hydrodynamic models flexible, cost-effective and accessible to non-specialists. By organising the available knowledge of estuarine modelling into an interactive and dynamic framework, the AI system provides the user with the necessary guidance and information for numerically solving hydro-environmental problems related to estuaries. As soon as a new problem is given to the system, the case-based module for estuarine modelling (CBEM) is activated. This module accesses information about estuarine models and estuaries to which numerical solutions have been previously applied. After comparison and evaluation, the case-based search engine returns from its memory the most effective modelling scheme available for the solution of the new problem. The system then calls the genetic algorithm (GA) module which optimises the physical parameters of the selected modelling procedure to suit the new application. The main focus of this paper is on the description of the GA module. This module is developed by combining the classical evolutionary approach with problem-specific information to carry out the required parameter optimisation. The effectiveness of this procedure is illustrated using a one–dimensional hydrodynamic model for the Upper Milford Haven estuary in UK. The comparison between manual and genetic algorithm based calibrations for this specific case suggests that the GA routine can very effectively calibrate estuarine models under realistic situations. This means a significant reduction in the time normally necessary for the implementation of a numerical modelling scheme.

1 INTRODUCTION

The use of computational modelling is essential for studying the complex behaviour of hydraulic systems. This is because of their efficiency and increasing ability to produce realistic and comprehensive simulations. However, modelling hydraulic phenomena needs to integrate different skills. In particular, in order to use these methods correctly it is necessary to have expertise in mathematical representation of the physical processes involved and numerical solution methods. Also a sound and exhaustive background in the hydrodynamic of natural water systems is usually needed [1]. Therefore, the expertise required for the application of numerical schemes in estuary modelling is relatively extensive and has become a limiting factor in restricting the use of these methods.

The need for models with intelligent user-interfaces has provided a stimulus for interdisciplinary research to integrate computational methods and artificial intelligence technologies. In this respect the emphasis has been mainly centred on developing information management systems based on data mining techniques such as artificial neural networks, genetic algorithms and knowledge discovery [2]. In particular, simulated evolution has been employed for estimating hydraulic parameters [3]. Recently, several applications based on genetic algorithms have been proposed for facilitating the solution of groundwater management problems [4], [5].

In this paper, an artificial intelligence framework for estuarine modelling, based on a combination of case-based reasoning and genetic algorithm technologies, is described. The physical behaviour of estuaries is intrinsically complex as they are the places where river and sea waters interact and mix. In addition, high concentration of human activities in these areas has resulted in an ever increasing risk of heavy pollution in estuarine systems. Numerical modelling is considered the most effective tool for simulating and predicting the effects of man made or natural changes within the estuaries. The aim of using artificial intelligence in assisting the design of numerical estuarine models is to enhance the effective exploitation of these techniques.

In the AI system developed in this work, the case-based reasoning module for estuarine modelling (CBEM) handles and evaluates the available knowledge on existing models and their applications. The genetic algorithm (GA) module, which becomes operational after CBEM stage, adjusts the selected solution to suit new situations. The details of CBEM module have been published previously [6] and will not be explained here. The main focus of this paper is hence on the GA module and its implementation. This module is created by the combination of the classical evolutionary approach and notions from the practice of the estuarine model calibration.

The developed GA module has been tested on a one-dimensional hydrodynamic model applied to the Upper Milford Haven estuary in Wales, UK. This model gives predictions about the water surface elevations and discharges under specific tidal conditions. The GA module provides an optimisation for this case.

---

[1] Chemical Engineering Department, Loughborough University, LE11 3TU, United Kingdom. Email: S.Passone@Lboro.ac.uk
[2] Computer Science Department, Loughborough University, LE11 3TU, United Kingdom. Email: P.W.H.Chung@Lboro.ac.uk

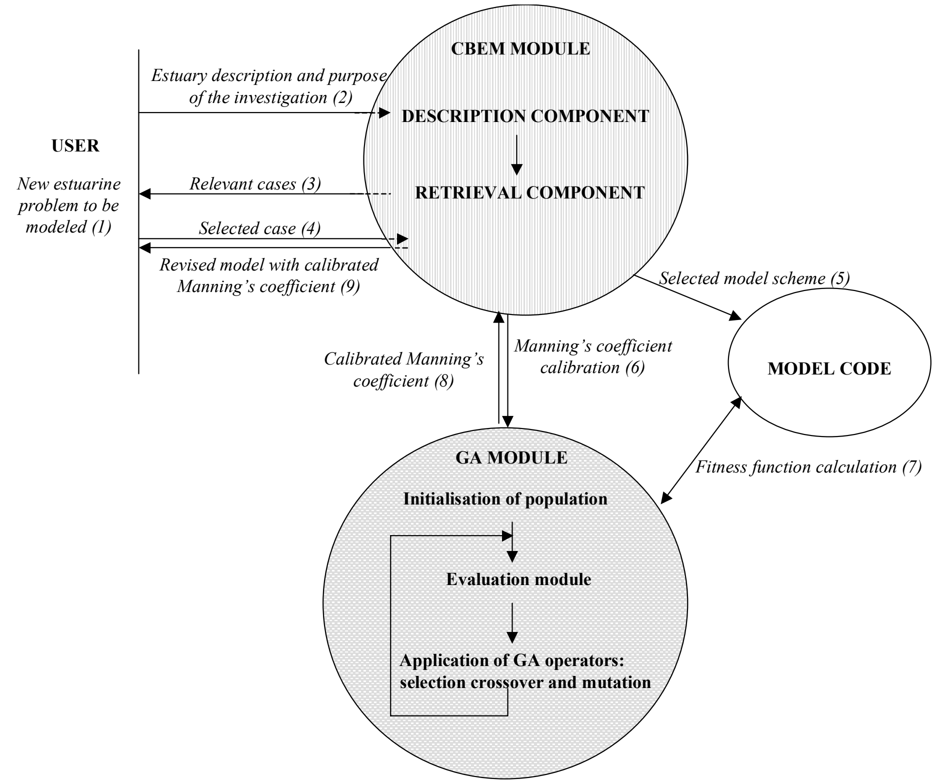


Figure 1. Architecture of the AI system for estuarine model design

The modelling experiments demonstrate the feasibility and efficacy of the GA based calibration in adapting a modelling scheme to the requirements of a new problem.

2 DESCRIPTION OF THE AI SYSTEM USED FOR ESTUARINE MODELLING

The AI system for estuarine modelling consists of two main parts: the CBEM and the GA modules as shown in figure 1. When the system is presented with an estuary to model, the new task is analyses through the main sub-processes of the case description, retrieval and adaptation [6]. Here the main attributes for a case are considered to be:

a) the modelling scheme suitable for a particular estuarine phenomenon

b) the configuration and the type of estuary to be modelled. Therefore, in the case description phase a new problem is categorised according to the physical characteristics of the estuary and the purpose of the investigation. The store of past cases provides information about the specific model schemes used for various types of estuaries. Comparing a new situation with the past experience the case retrieval selects and returns from its memory a modelling procedure which is able to give the best solution for the new case. The proposed model solution is then revised by the GA module for the new problem. The complexity of this job increases with the number of parameters. This is mainly due to the interdependency of physical parameters which affects model performance, even if only one parameter out of a large set is altered.

Calibration of numerical models for estuaries consists of determining the values of unmeasurable physical parameters to obtain the best agreement between the model simulation and the observed hydrodynamic behaviour of these water systems. After a model is calibrated it can be verified against a different set of input without altering the physical parameters. The calibration process is carried out either manually or using a computer program for numerical optimisation. Both manual adjustment of parameters and development of optimisation programs require expertise of an experienced modeller. The attraction of the GA module is that it can carry out this task independently by incorporating specific knowledge of estuarine model calibration and genetic algorithm operators.

After the GA based calibration, the user is presented with the most appropriate modelling procedure and physical parameters for a given problem. The information about the new case is stored in the case memory permitting the system to enhance its knowledge base.

3 THE GENETIC ALGORITHM

In the present system the GA module is specifically programmed to evolve appropriate Manning's coefficients for bottom friction in estuaries. The Manning's coefficient is a key parameter representing 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. The Manning's coefficient in estuaries typically varies within a range between 0.011 and 0.060 $m^{1/3}s$. In case of numerical models where the problem domain is divided into elements (fig. 2), the resistance to the flow within each discretised section is given by a specific value of the Manning's coefficient. Thus, calibrating the Manning's coefficient in a hydrodynamic model should result in a set of parameter values that gives a realistic simulation of the estuarine hydrodynamics.

3.1 Representation of chromosomes

The GA module is based on the utilisation of classical genetic algorithm theory. However, its application in the described optimisation scheme, which requires domain-specific knowledge, needs extensions beyond the established methodology. For instance, here the chromosomes are represented using the decimal base instead of the common binary alphabet. This provides the necessary correspondence between the set of Manning's coefficients and the chromosomes. Therefore, as the Manning's coefficient differs from one value to another only in the last two digits, the chromosomes are expressed as strings of integers corresponding to the second and third decimal places of the Manning's numbers (fig. 3). This representation is more practical since, to preserve the accuracy of modelling, the number of elements in a discretised domain is usually high (up to several hundreds) and increases with the dimensions of the flow domain. Therefore, with high 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 the GA operators.

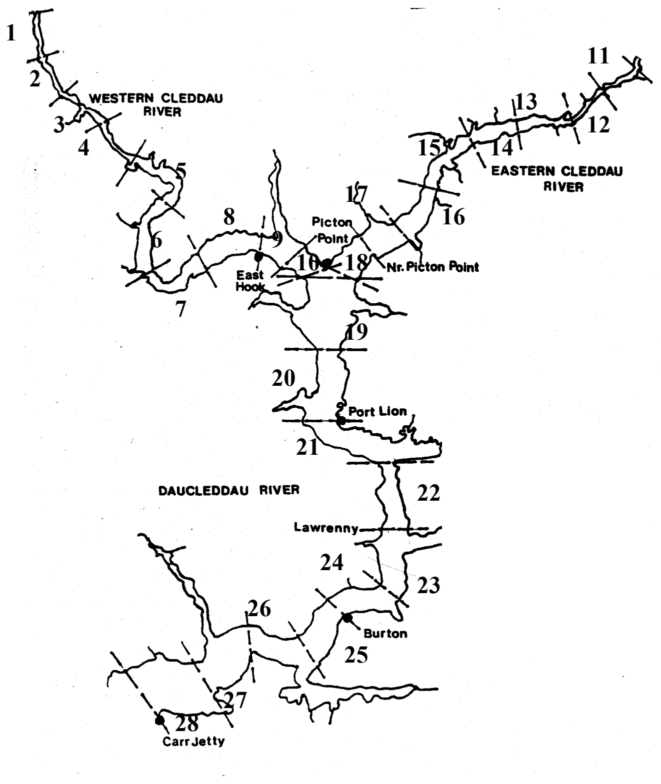


Figure 2. Discretisation of the Upper Milford Haven estuary, Wales, UK (a Manning's coefficient is assigned to each element in the domain)

MANNING'S COEFFICIENT SERIES

Element	n_1	n_2	n_3	n_4	n_5
Manning's number	0.020	0.020	0.021	0.034	0.021

CHROMOSOME REPRESENTATION

Genes	n_1	n_2	n_3	n_4	n_5
Allele	20	20	21	34	21

Figure 3. Example of a set of Manning's coefficients and representation of the corresponding chromosome

3.2 Generation of the initial population

Another feature of the present GA scheme is that it can use different modes for initialising the population of chromosomes. Although it is possible to generate the chromosomes randomly this is not a feasible choice because the Manning's coefficient remains the same, or varies very little, for adjacent elements in a discretised domain. The reason for this is that the bed resistance depends on the physical conditions of the estuarine location and reaches with similar physical characteristics are expected to have similar values for the Manning's coefficient. Therefore, using randomly generated coefficients there is the danger of obtaining unrealistic simulations. In order to have chromosomes that represent the estuarine flow resistance more closely, the GA program divides each chromosome into a number of segments corresponding to the zones of the estuary with specific physical characteristics. A value of the Manning's coefficient is then randomly generated for each part and assigned to the genes of the corresponding segment (zonation option) (fig. 4).

gene/element	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8
allele	35	35	28	28	28	28	20	20
	Estuarine zone A		Estuarine zone B				Estuarine zone C	

Figure 4. Example of chromosome initialised using the zonation option

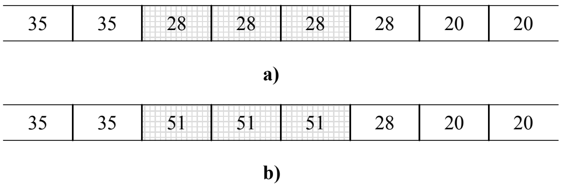
This flexibility can also be used to take into account the observation that the flow resistance generally decreases towards the estuary mouth. Based on this evidence, the GA program sorts the alleles in the initial population of chromosomes in a descending order, with lower values of the Manning's coefficient in elements towards the estuary mouth (scaling option).

The last feature implemented in the GA code for generating the initial population consists of seeding the cluster with appropriate Manning's coefficient series selected from the system's case-base. Based on the principle that similar problems should have similar solutions [7], for estuaries that do not significantly differ from one another, similar sets of Manning's coefficients are used to facilitate the GA search. However, only a limited number of pre-sets (less that 10%) are chosen from the whole population in order to ensure population diversity.

3.3 The GA operators

Starting from the initial population the subsequent generations are formed by selecting the chromosomes according to their fitness. The fitness criterion developed in this scheme is based on measuring the discrepancy between the simulated results obtained using a specific sequence and the corresponding field measurements. The sampling mechanism consists of the combination of roulette wheel and elitist approaches. The elitist logic stops the search to converge too quickly to unsuitable values. Therefore, this ensures that only around 10% of the population, which has the highest fitness values is transcribed in the next population.

The present scheme also contains different forms of the more common random mutation and crossover. The traditional genetic operators are modified according to the previously made observation of adjacent genes. The mutation routine implemented here operates by randomly changing the value of a gene and its closest neighbours. (fig. 5).



35	35	28	28	28	28	20	20

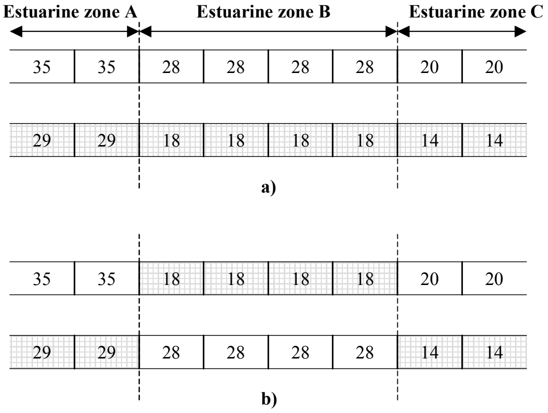
a)

35	35	51	51	51	28	20	20

b)

Figure 5. Mutation operator. a) Chromosome before the mutation; b) Chromosome after the mutation (the mutation transforms a gene and its closest neighbours)

The crossover operation consists of an exchange between the segments of two chromosomes associated with specific estuarine areas (fig. 6). The number of cut points can be more than one. At each crossover it is randomly chosen by the GA program, together with the segments that the chromosomes must reciprocate.



35	35	28	28	28	28	20	20
29	29	18	18	18	18	14	14

a)

35	35	18	18	18	18	20	20
29	29	28	28	28	28	14	14

b)

Figure 6. Crossover operator. a) Two chromosomes before the crossover; b) The offspring chromosomes after the crossover

4 THE FITNESS MEASURE

The calibration of a hydrodynamic model consists of finding an appropriate set of Manning's coefficients M_i, that minimises 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 $X_m=\{(x_m)^j_n, j=1,\ldots, J; n=1,\ldots,N\}$ and the set of all simulated values as $X_s=\{(x_s)^j_n, j=1,\ldots, J; n=1,\ldots, N\}$.

Hence, the discrepancy between X_m and X_s is given as [8]:

$$\rho(X_s, X_m)=\left[\sum_{j=1}^{J}\sum_{n=1}^{N}\left((x_s)_j^n-(x_m)_j^n\right)^2\right]^{1/2} \qquad (1)$$

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

$$\varphi_i=\frac{1}{\rho} \qquad (2)$$

In order to find for which chromosome equation (1) gives a minimum, water surface elevation for all of chromosomes in each generation must be simulated. Hence, x_s at each station j for the time levels n is calculated.

5 RESULTS AND DISCUSSION

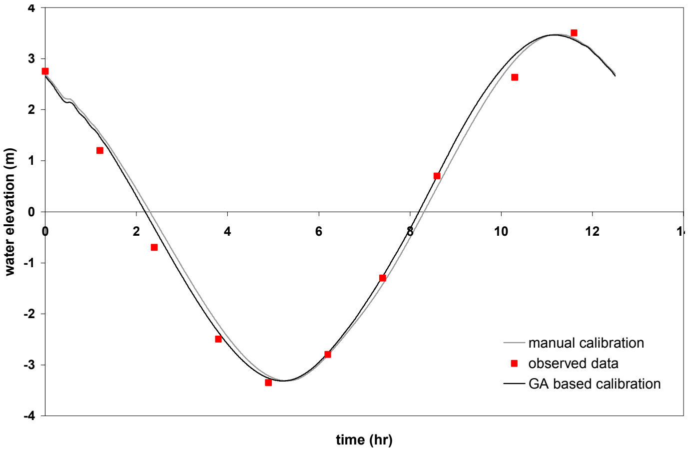
To test the described GA module the calibration of a one-dimensional Galerkin finite element hydrodynamic model of the Upper Milford Haven estuary is considered. The Upper Milford Haven estuary is situated in South-West Wales, UK. It is an example of "*ria*" estuaries, which are characterised by narrow and branching channels having a V-shaped cross-section and with low to moderate width to depth ratio [9]. This type of estuarine geomorphology means that the estuary can be adequately modelled using a one-dimensional approach. The present calculations are for a typical spring tide (25th April 1979), using the measured water surface elevations at Carr Jetty (fig. 2) and the fresh water inflow at the estuary head as boundary conditions. The estuary domain is discretised into 28 elements within which the equations of motion and continuity are solved to obtain water surface elevations at the nodal points of the elements (fig. 2). The detailed derivation of the mathematical model and the finite element solution scheme have been presented elsewhere [10] and are not repeated here.

For this specific case, the model calibration based on the described GA module is executed with the population set to 30 individuals and the rate of crossover and mutation equal to 0.5 and 0.01, respectively. The GA based calibration is carried out for 15 generations. The estuary is divided into three main zones corresponding to two branches (i.e.Western Cleddau and Eastern Cleddau rivers) and the main channel (the Decleddau reach) (fig. 2). Based on this partition of the estuary, the chromosome population is initialised using the zonation and the 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. This set of parameters was previously employed in the modelling of a similar ria estuary.

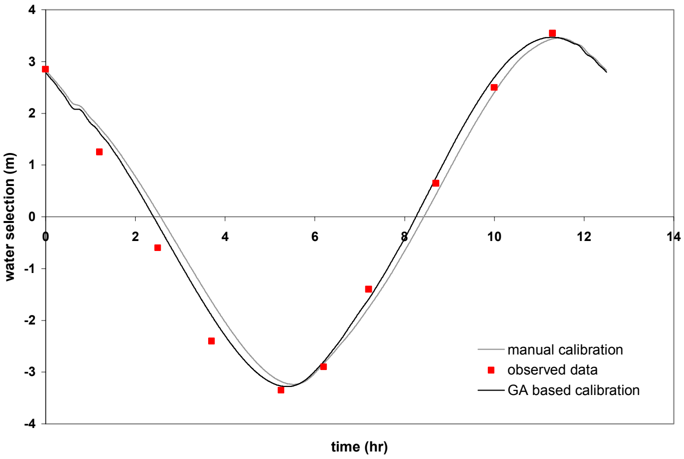
The calculation of the fitness function is based on the minimisation of the discrepancy between the model output and the observed data at Easthook and Picton Point (fig. 2). These locations are selected because the model is very sensitive to the hydrodynamic conditions at these two stations.

In figure 7 the simulated water surface elevations, generated using the set of Manning's coefficients selected by the GA module, are presented for Port Lion, Picton Point and Easthook situated, respectively, at 24, 26.8 and 28.2 km from the estuary mouth. These simulations at these locations show that the tidal wave has a regular shape within the main channel becoming gradually distorted as it propagates upstream from the junction into the branches. This is consistent with the geomorphological characteristics of the Upper Milford Haven estuary. The observed and simulated water surface elevations, obtained by the manual optimisation of the model, at the described locations are also shown in figure 7.

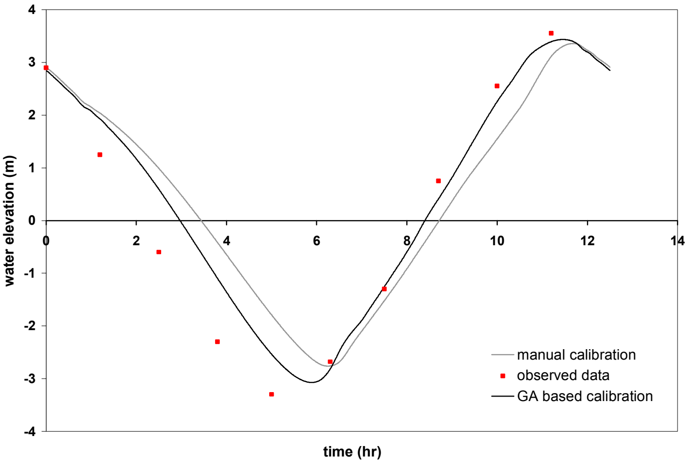
The comparison between the observed, manual and the GA based calibrations shows that the model using the GA module yields a better set of Manning's coefficients. The superiority of the GA based calibration over the trial and error optimisation is also demonstrated by considering the time necessary to carry out these two processes. In general, manual calibration of a model requires two weeks to one month (working time) while the GA based calibration takes 10 hours of CPU time in a shared SUN workstation.



a) Port Lion – 24 km from the estuary mouth



b) Picton Point – 26.8 km from the estuary mouth



c) Easthook – 28.2 km from the estuary mouth

Figure 7. Simulations of water surface elevations at a) Port Lion, b) Picton Point and c) Easthook

6 CONCLUSIONS

This paper presents an AI system for the design of estuary models. This system is based on the combination of principles of case-based reasoning and genetic algorithms. In particular, the use of evolutionary approach for calibrating hydrodynamic models is discussed. A robust methodology for optimising the unmeasurable friction coefficients in an estuary has been developed using the GA approach and discipline-specific knowledge. The classical genetic operations such as the initialisation of the population, crossover and mutation are modified to incorporate practical information available for estuarine model calibration. The preliminary results obtained for the Upper Milford Haven estuary indicate the feasibility and effectiveness of the GA based calibration under realistic conditions. The comparison with a trial and error calibration procedure demonstrates that the GA approach hybridised with a knowledge evaluation system can provide a superior optimisation process for complex hydro-environmental problems.

Additional tests are planned in order to verify the applicability of the present GA approach to other types of problems encountered in estuary modelling.

7 REFERENCES

[1] M.B. Abbott, *Hydroinformatics: Information Technology and the Aquatic Environment*, Avebury Technical, Aldershot, 1991.

[2] D.A. Savic and G.A. Walters, 'Hydroinformatics, data mining and maintenance of UK water networks', *Anti-Corrosion Methods and Materials*, **46**(6), 415-425, (1999).

[3] K. L. Prasad and A.K. Rastogi, 'Estimating net aquifer recharge and zonal hydraulic conductivity values for Mahi Right Bank Canal project area, India by genetic algorithm', *Journal of Hydrology*, **243**, 149-161, (2001).

[4] D.C. McKinney and M.D. Lin, 'Genetic Algorithm solution of groundwater-management models', *Water Resources Research*, **30**(6), 1897-1906, (1994).

[5] M.M. Aral, J.B. Guan, and M.L. Maslia, 'Identification of contaminant source location and release history in aquifers', *Journal of Hydrologic Engineering*, **6**(3), 225-234, (2001).

[6] S. Passone, P.W.H Chung, and V. Nassehi, *Development of a Decision Support System for Estuarine Modelling Using the Case-Based Reasoning Technology*, 16-24, Water Resources Management, WIT Press, Southampton, UK, 2001.

[7] S.J. Louis and J. Johnson, 'Solving similar problems using genetic algorithms and case-base memory', *Proceedings of the Seventh International Conference on Genetic Algorithms*, 101-127, Morgan Kauffman, San Mateo, CA, (1997).

[8] V. Babovic, Z. Wu, and L.C. Larsen, 'Calibrating hydrodynamics models by means of simulated evolution', *Proceedings of the First International Conference in Hydroinformatics*, A.A. Balkema, 193-200, (1994).

[9] K.R. Dyer, *Estuaries. A Physical Iintroduction*, John Wiley & Sons, Chichester, England, 2nd edn., 1997.

[10] J.H. Bikangaga and V. Nassehi, 'Application of computer modelling techniques to the determination of optimum effluent discharge policies in tidal water systems', *Water Research,* **29**(10), 2367-2375, (1995).