



This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

A Hybrid Reasoning System for Supporting Estuary Modelling

Sara Passone¹, Vahid Nassehi and Paul WH Chung²

Chemical Engineering Department, Loughborough University, Loughborough LE11 3TU

v.nassehi@lboro.ac.uk

Tel: +44 (0) 1509 222522

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 schemes 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 optimum parameters which can appropriately model the conditions and characteristics of any given estuary. Finally, the user is provided with a procedure that gives the best solution for a problem using the available hydrographic data and under the specified conditions. As an illustrative example and to show the applicability of the present CBEM system under realistic conditions a case study based on the simulation of salinity distribution in the Tay estuary (Scotland, UK) is given in this paper.

Key Words Case based reasoning, genetic algorithm, estuary modelling, IT systems

¹ Present address : Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA, 02139, USA

² 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 and Clifford, 2000). In addition, the human occupancy of coastal areas is meant to increase worldwide during the next 25 years (Hameedi, 1997). In particular, the degradation of estuaries and surrounding areas is expected to sharpen in developing countries where industrialisation is relatively recent and an effective legislation for waste dumping and wastewater discharges is still missing (Kennish, 2002).

Estuaries have been investigated by numerous disciplines such as geomorphology, ecology and hydrology for the past decades. Because of the complexity of such natural water systems, different investigations approaches have been adopted to identify the interplay among biologic, geologic, hydrologic, atmospheric, and chemical processes within estuaries. Oceanographers, engineers and natural scientists have been studying estuarine systems emphasizing on specific aspects over others. As a result, over 40 different definitions of estuaries have been given each of which is limited to a particular characterization of the estuary and its environment (Perillo, 1995). In addition, several estuary classifications have also been created each focusing on unique aspects of the estuarine environment such as geomorphology and physiography, tidal characteristics, hydrography and sedimentation within estuaries (Dyer, 1997). Although an integrated multidisciplinary research approach has been recently established, the sources of information and knowledge within the field of estuarine science still result scattered and little accessible.

Laboratory experiments, field measurements and numerical models are employed to understand and solve real estuarine problems. In particular, the application of computational hydraulics to estuaries has signified a real break-through in the attempt to fully interpret the mechanisms governing estuarine processes. Numerical models provide very accurate simulations with minimum time consumed (G. Thompson, 1993), revealing important aspects of estuarine environmental dynamics, which may not be evident from field measurements and analytical evaluations (French and Clifford, 2000). Furthermore, their use for investigating the hydrodynamics, sediments movement and mixing processes permits the exploration of multiple scenarios.

Although the currently used models are very sophisticated and characterized by a great rate of automation, the user of estuarine numerical models is required to have a high-level of expertise. The expert knowledge must be a combination of mathematics, physics, numerical methods and estuarine science (Dyke, 1996). The definition and study of any estuarine phenomenon via numerical modelling also relies on an example-by-example based knowledge. The user needs to be supported by his practical experience for the correct application of numerical models and interpretation of simulation results. Hampered by the shortcomings above described, traditional computational hydraulics has started developing in the last decade in a new direction with the intent of extending 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 (Abbot, 1991).

Different AI-hybrid systems that offer the user help and guidance at different operational levels of the decision-making process has been proposed for the management of water resources. Some of them handle uncertainty and risk management by using, for instance, the fuzzy set theory (Schulz and Huwe, 1997; Mpimpas *et al.*, 2001) or genetic algorithm technique (McKinney and Lin, 1994; Aral *et al.*, 2001) to define imprecise model parameters. Other focus on data capture, storage, processing and analysis to reduce the complexity of the available data and find new correlations and patterns (Abebe, 1999; Hall,

2002). In this perspective, AI techniques are used to create computer aided systems incorporating expertise and existing knowledge to guide and advise potential users. Within these knowledge management systems, modelling software is integrated with other information components to work as a single tool for the solution of water related problems (Cortes, 2001).

Numerical modelling directly benefits from its combination with AI techniques. Characterized by the organization 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 the user's goal and the problem domain's requirements (Knight and Petridis, 1992; Sophocleous and Ma, 1998; Chau and Chen, 2001). Following this approach an AI-hybrid system for supporting estuarine modelling, has also been developed. 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 and Chen, 2001), however, CBR technique provides more appropriate methodology in dealing with complicated problems arising in estuarine modelling. The examined knowledge, multidisciplinary and mainly supported by practical experience of real problems, is regarded as too difficult to be directly elicited from modelling experts. Instead, through the codification of previous studies and correct assessment of numerous assumptions for seen cases and problems the complexity of estuaries and the actual interaction of many problem factors can be easily estimated. Finally, CBR provides a good framework for integrating other 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 his experience to be employed under new circumstances, different types of knowledge techniques may be combined within the CBR framework to match the characteristics of the problem domain and comply with principles and assumptions of estuarine modelling theory.

The Case-Based Reasoning for Estuarine Modelling (CBEM) 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. The various knowledge related to the practice of estuarine modelling is made available in the description and retrieval components using the rule-based approach and the fuzzy set theory. For instance, within the retrieval component the model selection is also based on the established criteria of adequacy designed with respect to specific model properties. In addition, the implemented GA optimisation incorporates knowledge to guide the search. Such implementation is necessary to create a calibration under realistic situations as well as a significant reduction in the time normally required for the validation of a numerical model by other means.

The case study of Tay estuary is here presented to show that the adopted approach is convenient and effective for supporting the modelling of estuaries. The Tay estuary in Scotland, UK, is a quite meandering mesotidal-macrotidal watercourse with irregular width and depth. Hydrodynamically influenced by its geomorphology, the Tay estuary is classified as “*complex*” (Buck and Davidson, 1997). The Tay estuary is utilised throughout the entire paper to illustrate the system design and the implemented logic. In particular, CBEM is applied to define a model suitable to simulate the specific problem of salinity distribution for a management tool purpose.

GENERAL ARCHITECTURE Of The SYSTEM

CBEM consists of three main components: a case-based module, a genetic algorithm module and a library of numerical models (figure 1). These 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 purpose of the simulation. The separated descriptions for estuaries and models also permit identifying those aspects that contribute to define 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 readable to the user, who has to supply the necessary information, and the retrieval component speeding up the related process (Kolodner, 1993).

As the user decides to investigate a problem using CBEM, he/she enters the feature values of the estuary to be modelled in the estuary description scheme. He/she then defines 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 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 the co-operation 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 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 is assessed on the basis of the existing estuarine classifications for estuarine geomorphology, tidal characteristics and hydrography (water circulation, mixing process and stratification) (Dyer, 1997). By combining the information contained in these schemes, the estuary description is here organized 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 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 also include 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 classification focus on specific aspects of the estuarine environment and they 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 systematically organise 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 the fuzzy set theory, the width to the depth ratio and the degree of 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 and 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 part of cases already within the system's memory. The proposed scheme for model description (figure 3) was designed based on the 'Classification of the models of tidal waters' by Hinwood and Wallis (Hinwood and Wallis, 1975) and 'Guidelines of the use of computational models in coastal and estuarial studies' by Lawson and Gunn (Lawson and Gunn, 1996). The model is described in terms of: 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 correspondent model is retrieved by the system as being the most appropriate to the new problem.

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 the time required or the accuracy are accuracy, time required and simulation purpose.

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, for the case Tay estuary the user specifies "salinity distribution" as type of problem and the investigation purpose as "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 qualitatively (i.e. the degree of meandering, 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. 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),(b) and (c) and 0.75 to (d).
3. Determine the similarity values $S = \text{sim}(F_k^I, F_k^R)$, with sim as similarity function and, I and R referring to the input and retrieved cases, respectively.

4. Normalize 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 Tay estuary and the assessment of its similarity with the British estuaries of 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 3 cases of the initially retrieved 5 are in this list. The cases related to the Tees and Conwy estuaries are withdrawn from the set. In particular, these cases are eliminated because of the principle of exclusion, which tells: “If the estuary has inlets and at least one of the inlets is a “branch” then eliminate all 1-D models”.

Although the Fal estuary is not considered by CBEM as quite similar to the Tay estuary in terms of physical and hydrological characteristics (figure 6), however, the characteristics of its 1-D model network (i.e. “low” simulation time, “moderate” accuracy and “water quality” investigation purpose) (figure 7) make this specific case to be selected for investigating the salinity distribution problem for the Tay estuary with the purpose as “management tool”.

GA MODULE

For numerical models, case adaptation is essentially based on adjusting parameters to ensure that the model accurately simulates the real behaviour of the new case’s 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 the 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. The Manning's coefficient in estuaries typically varies within a range between 0.011 and 0.060 $\text{m}^{1/3}\text{s}$. For numerical models where the problem domain is discretised into elements (up to several hundreds), the resistance to the flow is expressed by associating to each section a specific value of the Manning's coefficient. Thus, calibrating the 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 the 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 and 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).

GA module carried out the calibration of the Manning's coefficient independently by combining the classical evolutionary 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 the 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 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 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 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 resistance to the flow changes with respect to the variation of the estuarine physical characteristics and reaches with similar physical characteristics are expected to have similar values for the Manning's coefficient. Therefore, it is expected adjacent elements to have similar values for the 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 the 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 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 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 and Johnson, 1997), estuaries that do not significantly differ from one another should have similar sets of Manning's coefficients. The sets are preventively 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 $H_m = \{(h_m)_n^j, j=1, \dots, J; n=1, \dots, N\}$ and the set of all simulated values as $H_s = \{(h_s)_n^j, j=1, \dots, J; n=1, \dots, N\}$.

Hence, the discrepancy between H_m and H_s is given as:

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

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

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

In order to find which chromosome gives a minimum for equation (1), 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 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 in the population. This stops the search to converge 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 the 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's segments corresponding to specific estuary's zones. The crossover operator swaps between chromosomes segments which correspond to specific estuary's 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. Thus, the chromosomes are mutated by changing the value of a randomly chosen gene and its closest neighbours (figure 9).

GA calibration for The Tay Estuary

Once the 1-D network model has been selected for being applied to simulating the salinity distribution problem into 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 to the particular hydrodynamics and geomorphological characteristics of the Tay estuary. The present calculations are for a typical spring tide (12th June 1972), using the measured water surface elevations at 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 individual 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 the 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 into the Fal estuary using the 1-D 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 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 the 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, the 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 yields a better performance. In particular, while the two sets generate similar water elevations at Newport, the GA based calibration better copes 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.

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. (say which model)

CONCLUSIONS

The described 'hydroinformatics' system is shown to provide 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 appropriate management decisions. It is also shown that using module based algorithms lengthy operations such as 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.

REFERENCES

1. Abbott, M.B. (1991) *Hydroinformatics – Information Technology and the Aquatic Environment*, Avebury Technical, Aldershot
2. Aral M.M., Guan J.B., Maslia M.L. (2001) 'Identification of contaminant source location and release history in aquifers' *Journal of Hydrologic Engineering* **6**(3): 225-234
3. Buck, A.L. and Davidson, A.C. (1997) *An Inventory of UK Estuaries*, Joint Natural Conservation Committee, Peterborough
4. Chau, K.W. and Chen, W. (2001) 'A Fifth Generation Numerical Modelling System in Coastal Zone', *Applied mathematical Modelling*, **25**, 887-900
5. Chung, P.W.H. and Inder, R. (1992) 'Handling Uncertainty in Accessing Petroleum Exploration Data', *Revue de L'Institut Francais du Petrole*, **47**(3), 305-314
6. Cortés U., Sánchez-Marrè, M., Sangüesa, R., Comas, J., Roda, J.R., Poch, M. and Riaño, D. (2001) 'Knowledge Management in Environmental Decision Support System', *Ai Communication*, **14**, 3-12
7. Dyer, K.R (1997) *Estuaries: a Physical Introduction*, John Wiley & Sons, New York
8. Dyke, P. (1996) *Modelling Marine Processes*, Prentice-Hall, London
9. French, J.R. and Clifford, N.J. (2000) 'Hydrodynamic Modelling as a Basis for Explaining Estuarine Environmental Dynamics: Some Computational and Methodological Issues', *Hydrological Process* **14**, 2089-2108
10. GESAMP (1990) *The State of the Marine Environment*, published by United Nations Environment Programme

11. Goldberg, D.E.(1989) Genetic Algorithm in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, Massachussetts, U.S.A.
12. Grefenstette, J. J. (1987) 'Incorporating Problem Specific Knowledge into Genetic Algorithms', In Genetic Algorithms and Simulated Annealing, L. D. Davis (Ed.), 42-61, London: Pitman
13. Hall, J.W. (2002) 'Random Set Uncertainty Handling in Hydroinformatics', Proceedings of the 5th International Conference on Hydroinformatics, edited by I.D. Cluckie, D. Han, J.P. Davis and S. Heslop, 1388-1393
14. Hammedi, 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, Rotterdam, the Netherlands: Kluwer Academic Publishers
15. Hinwood, J.B. and Wallis, I.G. (1975) 'Classification of Models of Tidal Waters', Journal of Hydraulics Division, Proc. of A.S.C.E., 101, HY10 1315-1351
16. Kennish, M.J. (2002) 'Environmental Threats and Environmental Future of Estuaries', Environmental Conservation 29(1) 78-107
17. Knight B., Petridis M. (1992) Flowes - An Intelligent Computational Fluid-Dynamics System Engineering Applications of Artificial Intelligence 5(1): 51-58
18. Kolodner, J. (1993) Case-Based Reasoning, Morgan Kaufmann, San Mateo, California
19. Lawson, J. and Gunn, I.J. (1996) Guidelines for the Use of Computational Model in Coastal and Estuarial Studies, Report SR 450, HR Wallingford, Oxon, UK
20. McDowell, D.M. and O'Connor, B.A. (1977) Hydraulic Behaviour of Estuaries, Macmillan Press LTD, London
21. McKinney, D.C. and Lin, M.D. (1994) 'Genetic Algorithm Solution of Groundwater-Management Models', Water Resources Research, 30(6) 1897-1906

22. Mpimpas H., Agnastopoulos P. and Ganoulis, J. (2001) 'Modelling of Water Pollution in the Thermaikos Gulf with Fuzzy Parameters', *Ecological Modelling* 142(1-2), 91-104.
23. Perillo, G.M.E. (1995) *Geomorphology and Sedimentology of Estuaries*, Elsevier
24. Schulz K. and Huwe B. (1997) 'Water Flow Modeling in the Unsaturated Zone with Imprecise Parameters Using a Fuzzy Approach' *Journal Hydrology* **201**(1-4), 211-229
25. Sophocleous., M. and 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
26. Thompson, G. (1993) 'Mathematical-Models and Engineering Design', *Journal of the Institution of Water and Environmental Management*, 7, 18-23

Figure 1

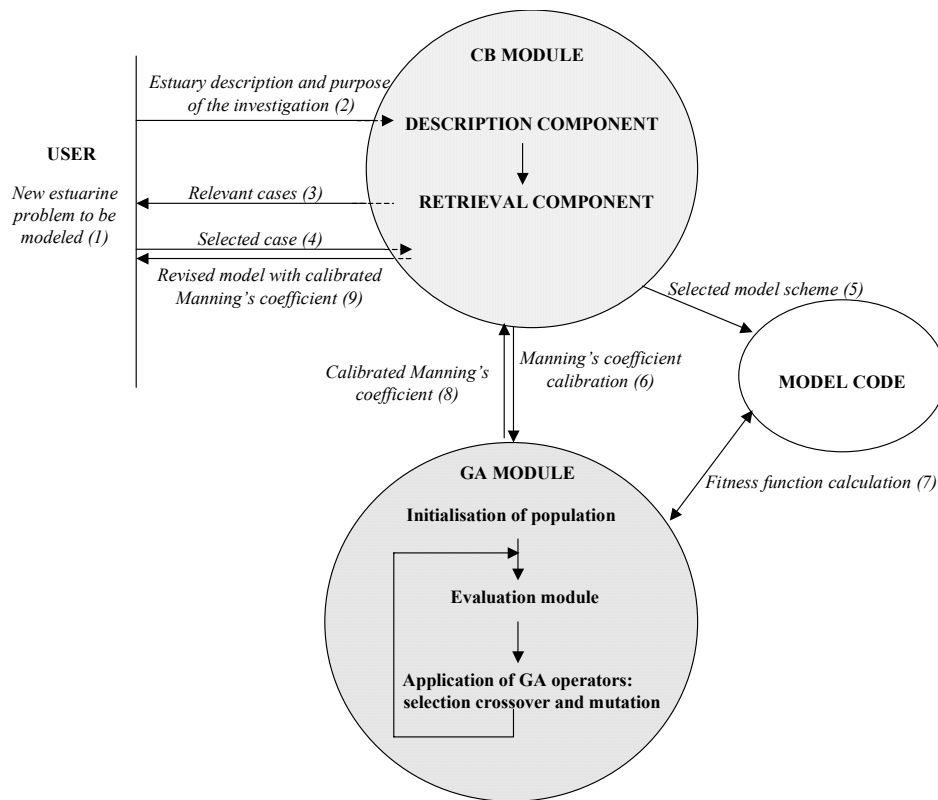


Figure 2

The screenshot shows a Microsoft Access form titled "ESTUARY" with the following data:

Physical characteristics			
name	TAY	ID	4
country	Scotland	latitude	55
geomorphological type	complex	tidal range	mesotidal/macrotidal
total area (ha)	12128	intertidal area (ha)	5583
max channel bank area (ha)	2600	channel length (km)	54
valley length (km)	33	meandering/sinuosity	meandering
min. width (m)	2000	max. width (m)	5000
bed	flat bed	avr. width (m)	3000
avr. depth (m)	10	avr. width to depth ratio	very high

Hydrographic characteristics			
tot. freshwater inflow (m3/s)	260	tidal volume(m3)	63500000
salinity	vertically mixed	salinity limit (km)	45
averaged velocity (m/s)	2	tidal period (hr)	12.5
<input checked="" type="checkbox"/> inlets	How many inlets? 2	<input type="checkbox"/> coriolis effect	<input type="checkbox"/> wind

comments: generally meandering channel

Buttons: Apply Model..., First, Previous, Next, Last, New

Record: 4 of 8

Description of the Tay estuary, Scotland, UK

Figure 3

The screenshot shows a software window titled "frmMaintModels" with a "MODELS" form. The form contains several fields and checkboxes:

- estuary:** TAY
- problem:** salt intrusion
- dimension:** 2-D avrg on depth
- numerical method:** finite element
- purpose:** management tool
- accuracy:** moderate
- manning's coefficient:** varying manning
- dispersion coefficient:** salt interpolation

Governing Assumptions:

- wind
- coriolis force
- branching
- slope

limitation of use: [Empty text box]

time consumed: high

Record: 4 of 12

Figure 4

The image shows a software dialog box titled "frmselectioncriteria" with a close button in the top right corner. The main heading is "Model Selection".

Fields and controls include:

- Estuary:** A text input field containing "TAY".
- problem:** A dropdown menu with "salinity distribution" selected.
- purpose:** A dropdown menu with "management tool" selected.
- cut-off:** A text input field containing "0.5".
- Approximation:** A section with four checkboxes:
 - coriolis effect
 - wind effect
 - branching
 - bed slope

At the bottom, there are two buttons: "OK" and "Cancel".

Figure 5

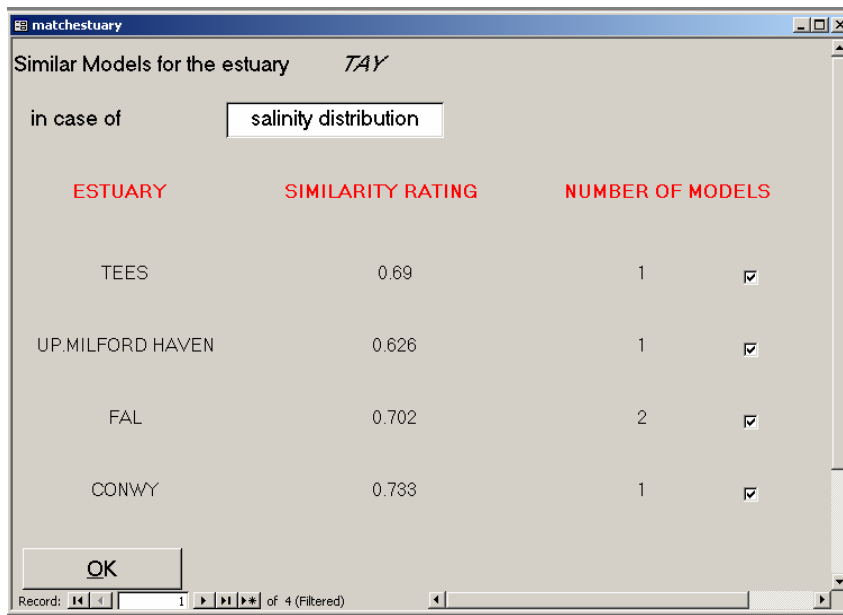


Figure 6

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 8

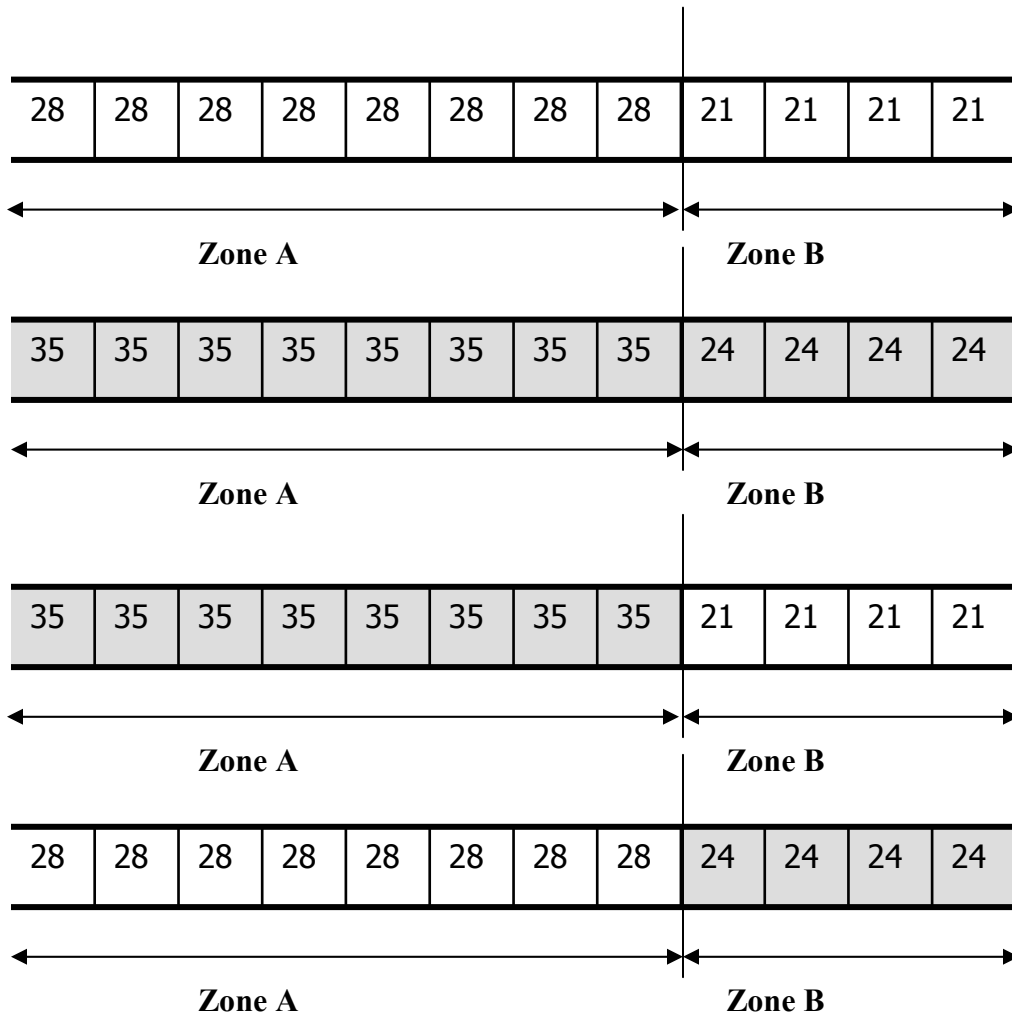


Figure 9

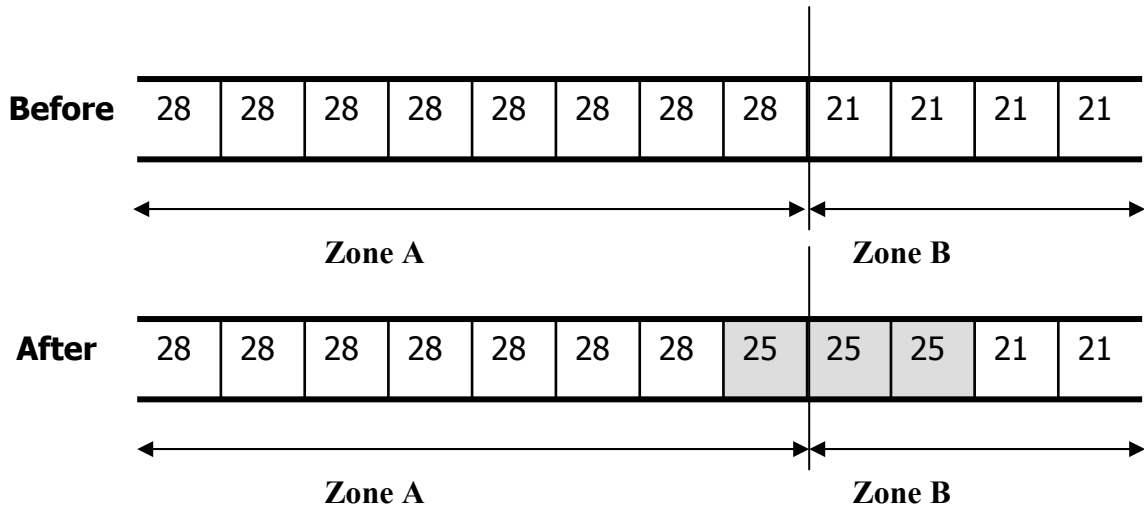


Figure 10

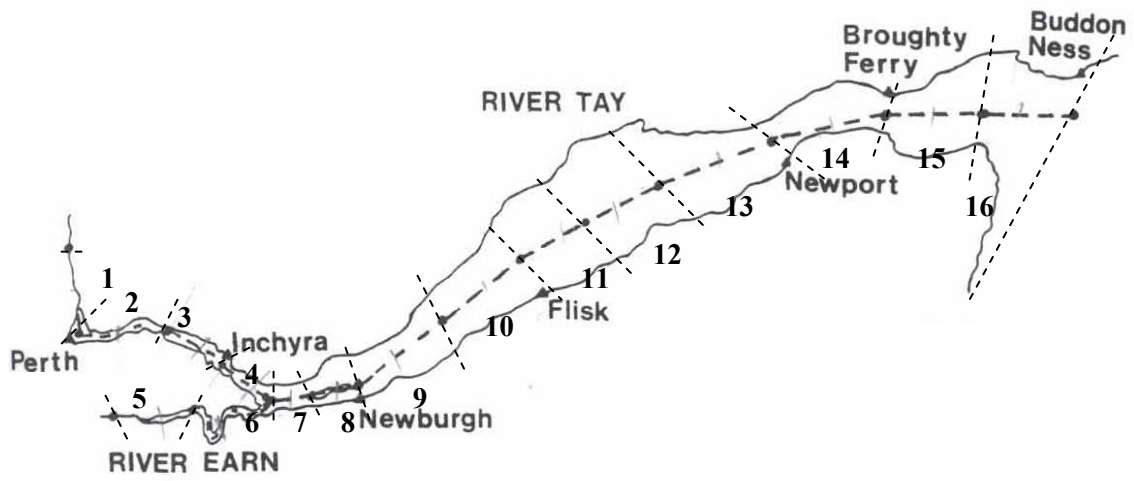
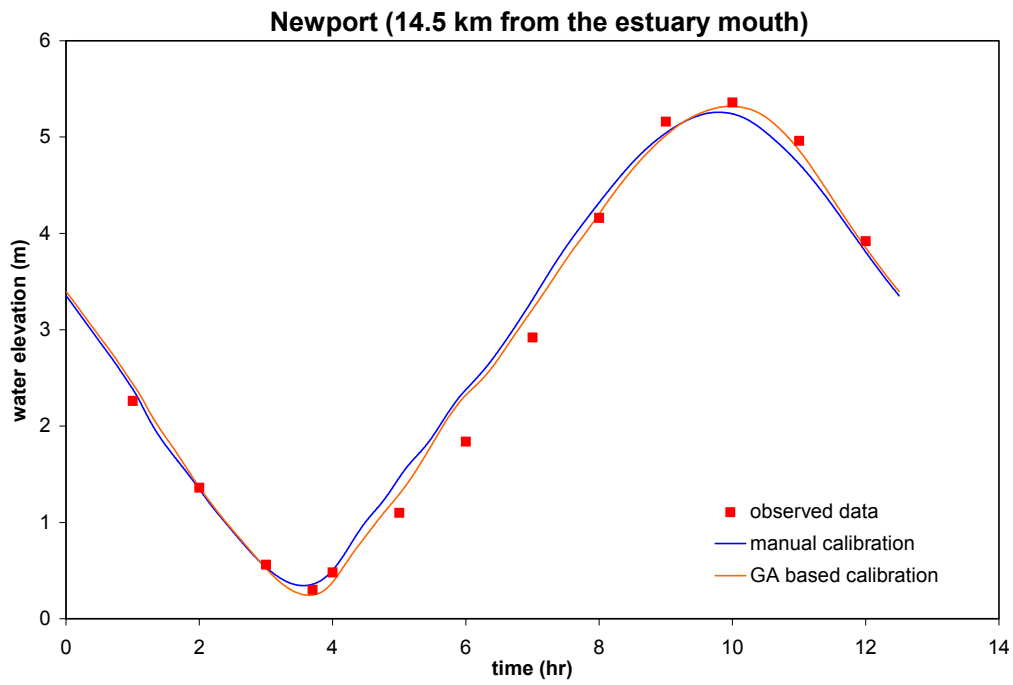
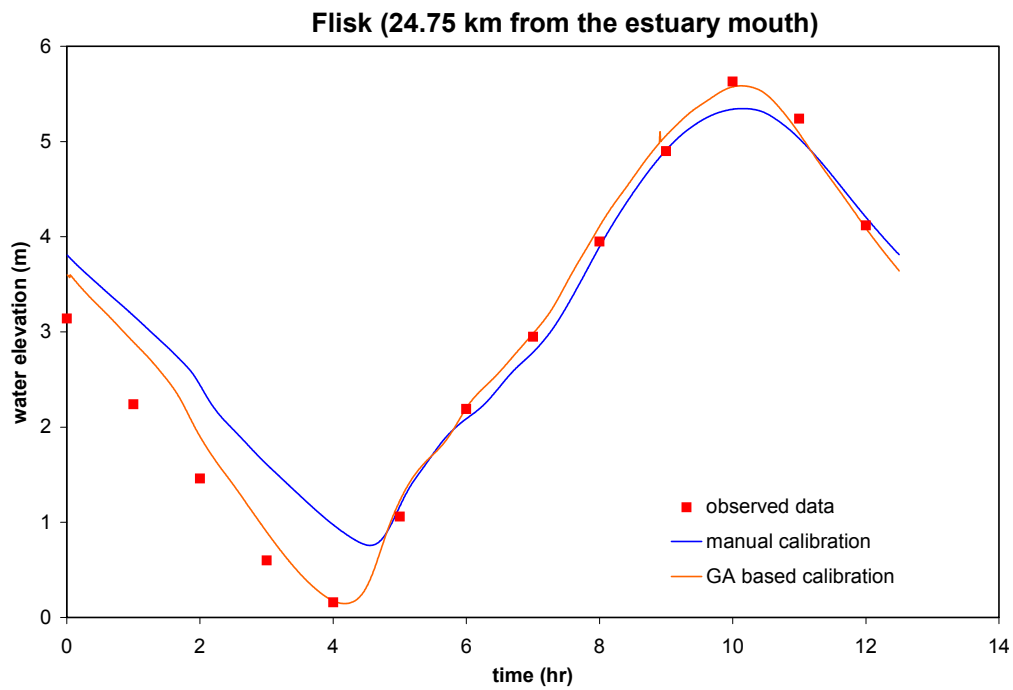


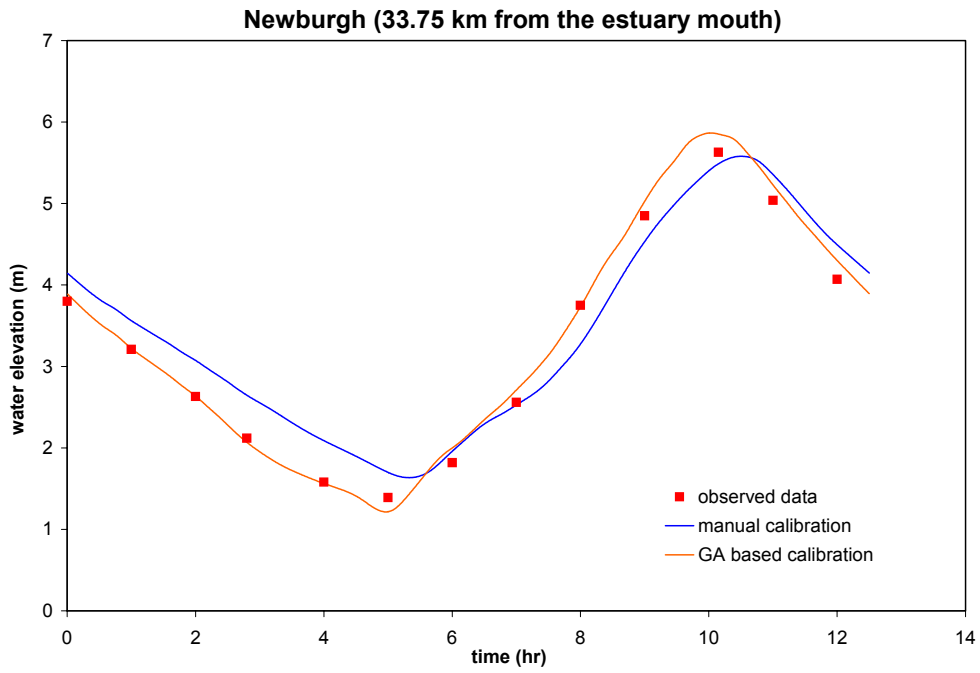
Figure 11



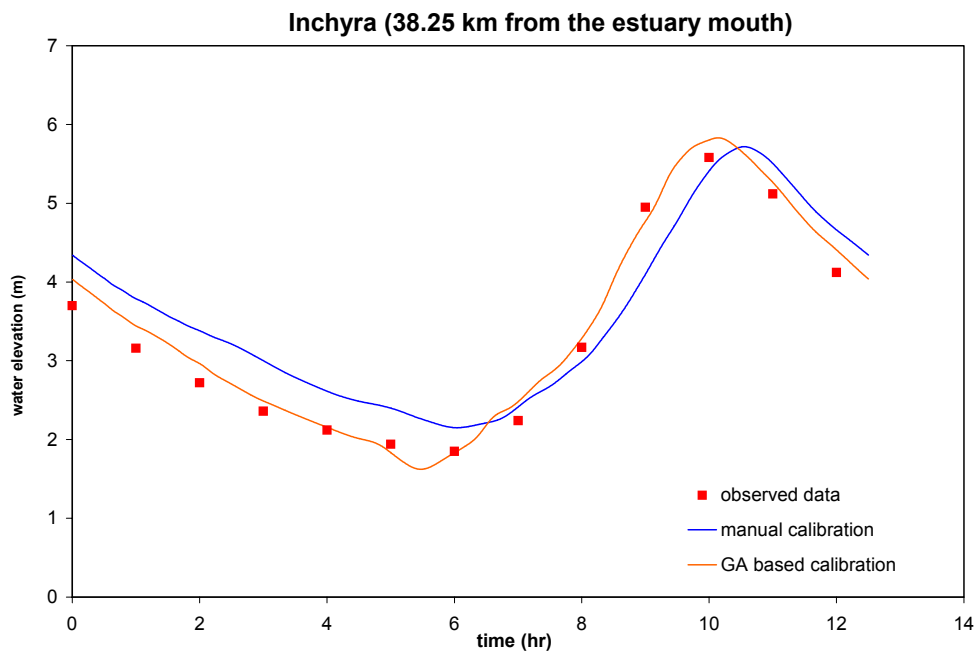
a)



b)



c)



d)