Towards the hydraulics of the hydroinformatics era
L'hydraulique à l'ère de l'hydroinformatique

M.B. ABBOTT, Knowledge Engineering, *Avenue François Folie 28, B 28, 1180 Brussels, Belgium*

V.M. BABOVIC, DHI, *Agern Alle 5/11, 2970 Hørsholm, Denmark*

J.A. CUNGE, IHE, *Westvest 7, P.O.Box 3015, 2601 DA, The Netherlands*

SUMMARY

Hydroinformatics is the study of the flows of knowledge and data related to the flow of water and all that it transports, together with interactions with both natural and man-made, or artificial, environments (Abbott, 1991). Hydraulics, understood as the study of flows of water, more recently extended to include the transport of matter in all its forms with these flows, is accordingly central to hydroinformatics. Without hydraulics, no hydroinformatics!
From this situation it may at first appear as though hydroinformatics provides only a new periphery to hydraulics: a new way of transmitting hydraulics knowledge and data to society. In practice, however, the way in which hydraulics is viewed and practised is itself now changing as a result of its incorporation into the new paradigm that hydroinformatics provides. The first purpose of the present paper is to introduce some of the changes that are currently proceeding in hydraulics under the influence of developments occurring in hydroinformatics. The second purpose is to indicate the consequences of these changes for the application of hydraulics within society, and thus for the future direction of hydraulics and hydroinformatics themselves.

RÉSUMÉ

L'hydroinformatique est l'étude des flux de connaissances et de données concernant l'écoulement de l'eau et de tout ce qu'elle transporte, ainsi que les interactions avec les environnements naturels ou artificiels (Abbott 1991). L'hydraulique, comprise comme l'étude des écoulements d'eau, plus récemment étendue au transport de matière sous toutes ses formes, dans ces écoulements, est en conséquence au centre de l'hydroinformatique. Sans hydraulique, pas d'hydroinformatique!
A partir de là, tout se passe comme si l'hydroinformatique constituait seulement un nouveau domaine périphérique de l'hydraulique, une nouvelle façon de transmettre à la société le savoir et les données hydrauliques. Dans la pratique, cependant, la manière dont l'hydraulique est vue et pratiquée est elle-même en train de changer, du fait de son insertion dans les nouveaux paradigmes que fournit l'hydroinformatique. Le premier objectif du présent article est d'introduire quelques uns de ces changements qui apparaissent couramment en hydraulique sous l'influence des développements de l'hydroinformatique. Le second objectif est d'indiquer les conséquences de ces changements pour les applications de l'hydraulique dans la société, et donc pour la direction future de l'hydraulique et de l'hydroinformatique elles-mêmes.

The hydraulic engineer in the post-symbolic era

The first and most obvious change that has occurred in hydraulic engineering is in the way in which the hydraulic engineer works. Like most other engineers nowadays, the hydraulic engineer works for a large part, and in many cases for the most part, through the graphical user interface of a computer. The era in which the engineer worked with symbols, making calculations directly from equations and the curves of graphs of such equations, is mostly over. Today, the engineer works for the most part with signs. Whereas the symbols of an earlier era *replaced* the world in the mind of the engineer, at least while he or she manipulated these symbols, the signs of the graphical user interface (the buttons, the pull-down menus, etc) *point towards* the world in the mind of the engineer. More basically, while in the earlier era the engineer was a repository of knowledge made expressible in symbols, that is, a *knower*, the engineer is now primarily a repository of the sum of all the means to access knowledge, so that he or she is primarily a *consumer of knowledge* made expressible in signs (Baudrillard, 1963; Lyotard, 1979//1982 and 1998, Appignanesi and Garatt, 1995). Correspondingly, the device that was previously a *computer*, as a means of making computations, now becomes a *knowledge processor*, as a means of manipulating encapsulated knowledge. Similarly, what was previously a *data network*, as a means merely of accessing data, now becomes an intranet, or even an extranet, or more generally an *internet* (in the generic sense and so with a lower-case 'i') as a device for communicating knowledge in the first place and data only in the second place. The new era, in which the engineer no longer works with symbols in the capacity of a knower, but instead works with signs in the capacity of a consumer of knowledge, is called quite generally the post-symbolic era (Abbott, 1998 a and b; Jonoski and Abbott, 1998; and, especially, Abbott, 1999 b).

Tool builders and tool users

Corresponding to this change in the status of engineer, and indeed in the status of all knowledge users, a division opens up between those who encapsulate knowledge for consumption, on the one hand, and, on the other hand, those who access and use this knowledge. The equipment whereby knowledge is transformed and transferred belongs to the category of tools, so that we identify here a division between *toolmakers* and *tool users*. Correspondingly the most immediate and direct consequence of this division is that the most advanced hydraulics knowledge has become useable to far more persons than hitherto. From their inception in 1986, fourth generation modelling systems alone have increased the number of persons who are able to draw upon and use

---

Revision received June 28, 2000. Open for discussion till February 28, 2002.

this knowledge by an order of magnitude every five years. As of 2000, well over 6.000 organisations were using tools of this class, representing around 15.000 users at any one time in more than 100 countries. This development has in its turn only been made possible through the radical changes that have occurred within modelling tools of this class and through the new social-institutional arrangements that have been set in place for their dissemination and support. This first stage in the more widespread application of hydraulics knowledge is illustrated in Fig 1 together with the later stages which will be introduced later in this paper. In principle, a numerical-hydraulic modelling tool is one that encapsulates generic hydraulic and related knowledge in such way that this may operate on site-specific data in order to provide site-specific knowledge. The extent to which this knowledge comes to presence in the mind of the tool user naturally depends upon the ability of that user to interpret the (mostly graphical) output of the system, and so upon that user's particular *knowledge frame* (Abbott, 1993). A considerable part of the effort put into fourth generation modelling systems has been directed to increasing the rate of knowledge processing by their users by the provision of more appropriate interfaces and supporting facilities, while most of the rest of this effort has been expended upon extending the range of application of these systems and automating their instantiations. Thus, whereas in 1985 some eighty percent of investment was still associated with the numerics of modelling tools, by 1998 at least seventy percent was being used for enhancing knowledge transmission facilities. The emphasis has passed correspondingly from number organisation and application, or *numerics,* to sign production and organisation, or *semiotics.* It has thus passed over to the study of sign production, distribution and consumption, which belongs to the subject area of the *theory of semiotics* (e.g. Eco, 1976; Klinkenberg, 1996). In the simplest terms, we observe the transformation:

Numerics → semiotics

This transformation is naturally reflected in the composition of the design and production teams of the tool producers, where quite new skills have had to be introduced alongside the existing skills. The highly visible developments in the latest versions of modelling tools – common-run-time environments, installation and license systems, hypertext-extended help facilities, seamless coupling to standard GIS and CAD packages and data management systems, video and multimedia presentation environments, etc. – are themselves enabled by a wide variety of hidden features, such as utility file standards, including data file standards and parameter file standards for text files. These developments are supported in their turn by a variety of computer-aided software engineering (CASE) tools and other advances in software engineering, continuing into such advances in communication technologies as those that are enabled by XML, ActiveX/Com, Java and CGI technologies. This is, however, only the beginning of a process that has continued to accelerate ever more rapidly as hydroinformatics has progressed beyond fourth-generation modelling, as is also introduced towards the end of this paper.

Experiences with the applications of fourth-generation systems

The very great success of modelling and simulation tools working within the post-modern paradigm of knowledge consumption in this area naturally also brings many problems with it. Although

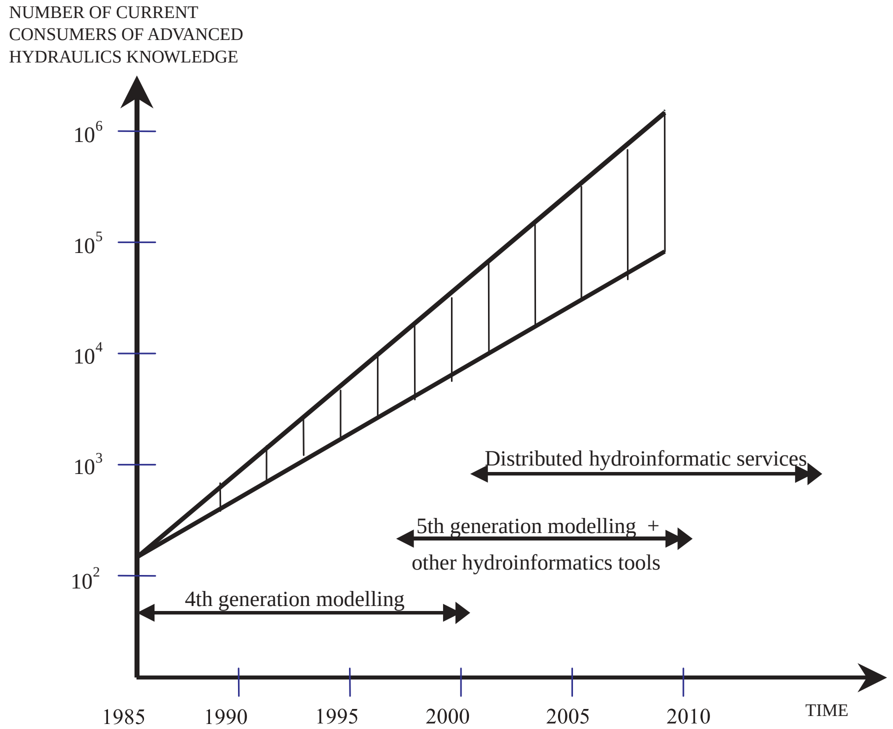

Fig. 1.Historical and predicted growths in the numbers of consumers of high-level knowledge in hydraulics, hydrology and water resources

not strictly speaking 'new', these problems have become exacerbated by inadequacies in the knowledge frames of many of the users of such systems. The most notable class of problems – and one that in fact dates back to the era of physical-hydraulic models – is associated with the process usually known as 'calibration', and this will be used here in order to exemplify the more general problem. It is well known how the results of models of fluid flow can nearly always be fitted to measurements of hydraulic behaviour recorded in a few places by varying the values of certain parameters, and most commonly the roughness coefficients, in the model. This process can however easily hide gross inaccuracies in the description of the modelled domain, such as in topography, in local head losses and in the dimensions of structures. On the other hand, the range of variation of, for example, Manning coefficient values is rather well known. In practice, it is almost certainly safer to make an error in estimating these values on the basis of a visual inspection of the terrain than to try to 'calibrate' a model by attributing unrealistic values to these coefficients in an attempt to compensate for unknown information or, worse, for physical phenomena not represented by the model laws, as expressed in the equations used in the models. In the former case one knows the range and the interval of possible imprecision. The latter is truly a black box. It does not mean that this kind of calibration should be forbidden altogether, but is should take place, within this example, only within a very strictly, physically-known range of Manning coefficient values and only where the type of bed or flood plain cover is also well known.

In general, it is acceptable in practice to calibrate roughnesses when small differences within a known range of variations cause, for example, computed and observed hydrographs to coincide. It is absurd however to do the same when differences cannot be explained or reduced by varying known and experimentally established coefficients, because this means that one has 'forgotten' something, such as singular head loss (which in fact provides the least disadvantage when one is trying to replace its influence by a longitudinally-distributed Manning coefficient of an unlikely value). Only too often the 'forgotten' part is a missing physical law or a process. In the former, acceptable situation, the model will be predictive even without calibration. In the latter, unacceptable situation, it will not be predictive even with an apparently excellent calibration.

Illustrative examples

A few examples, selected because they are simple and because they all correspond to cases which have actually occurred, may serve to illustrate the above thesis on the role of calibration in good modelling practice.

Example 1: Calibrating models within the context of a decision-making problem

A very detailed unsteady flow model was built for a several-hundred- kilometre long reach of a topographically very-well-documented European river. Water stages during recent floods with discharges culminating at some 3500 m^3/s were systematically recorded, as is usually the case, only at a limited number of stations, in this case some 20 kilometres apart. The model was carefully calibrated by varying the roughness coefficients in order to obtain the best possible coincidence between the computed hydrographs and those recorded at the stations. Higher floods occurred in the past (some 100 to 150 years previously) but their influences were not properly recorded and, moreover, the river bed had since been seriously modified, not only by natural processes but also by dikes. The dikes were intended to protect the valley from catastrophic discharges so long as these discharges were lower than about 7000 m^3/s. The decision-making process was concerned with crisis management. If the flood forecast gives a certain discharge propagating along the river, such that water may be expected to flow over the dikes, the authorities should have some 24 hours in which to evacuate the population and transportable assets from the threatened areas. Costs and damages resulting from a wrong decision are very considerable and, in particular, would still be considerable in the case of a false alarm. A flood peak of some 7000 m^3/s was accordingly simulated with the calibrated model. Maximum computed water stages were then found to be some 15 cm lower than the dike crests. The calibration of the model having been based upon the difference between observed and computed hydrographs for a 3500 m^3/s flood being within 10 cm, the authorities could have concluded that for an upstream 7000 $m^3/_S$ flood there was no reason to evacuate the population. Many hydraulic engineers, however, would conclude otherwise. They would do so simply because the model was calibrated with respect to water stage recording stations some 20 kilometres apart. Because of this, the calibrated roughness coefficients had to compensate for a considerable range of missing hydraulic and topographical data, with all manner of errors and assumption being implicitly made: The ranges of variation in all these phenomena were simply not known. It was pointless to run the usual 'sensitive range computations', corresponding to changing all Manning n values by +/–15%, because the real life situation may well have been very different from the resulting 'black box'-encapsulated 'reality' and flows under such circumstances may not at all necessarily vary linearly with *calibrated n* values. The predictive capacity of the model for this specific decision-making purpose was accordingly dubious, and the social consequences of its use could have been catastrophic.

Example 2: celerity model calibration and topographical data

This difficulty was encountered many more years ago when modelling a large South American river. The topography of the river was supplied by the client for whom a surveying company had done the work. The flood, simulated with one-dimensional modelling software, propagated much quicker in the model than in reality. There was no way to obtain a satisfactory coincidence between computed and observed hydrographs except by assuming values of Manning n values which could only possibly be justified if the flood plain was covered by jungle. This was quite simply impossible at that latitude. It was nevertheless quite possible by using these values to deliver a calibrated model to the client, whose requirements were limited to reasonably coinciding hydrographs. What actually happened was that the valley flood plain was generally very wide with a number of narrower locations. The surveyor made large savings by carrying out valley flood plain cross-section surveys only in the narrow locations. Thus the width of the modelled cross-sections was taken as much smaller than it was on average in reality and the computed celerity was seriously affected accordingly. The predictivity of the model for higher discharges was also jeopardised because of wrongly simulated storage volumes. However, the model had been 'calibrated'. Once the reason for the discrepancies had been found and a realistic topography introduced, calibration led to very limited changes in the roughness coefficient values that could have been estimated from the bed material and land cover description in the first place.

Example 3: Backwater calibration problem and topographical data

A similar problem was experienced recently when using a two-dimensional model to calculate a steady flow in a river for which observed backwater curves were available. The observed and computed backwater curves did not fit. Here again the data supplier (this time the client) was unaware of the mechanism and requirements of the model. He decided by himself that the available data (on cross-sectional geometry) were not all necessary and that the use of all the available cross-sections would unduly increase the cost of the modelling, of the data processing, and even of computer time! Hence he supplied the modeller with half of the available cross-sections, simply eliminating every second one. It was *possible* to calibrate the Manning coefficients in the model and obtain the required coincidence. But what predictivity could such a model have? Actually there was no need to calibrate at all: once all the cross-sections had been taken into account, the coincidence was very good when using 'standard' coefficients without calibration. Playing with the data was even more deleterious to realism in this example because two-dimensional modelling was used.

Prediction without calibration

The thesis that presents itself here is that deterministic models which are based on a set of laws representing all features that are essential and of interest to the problem should, in principle, *not* need to be calibrated at all. In most cases, calibration is based on insufficient data and justified by an imaginary and often false concept of improvement, and this malpractice usually only reduces the model's predictive capacities and usefulness for engineering purposes. However, this thesis is not sustainable unless at the same time one is sure that the equations or other representations of laws that one is using do in fact correspond to the behaviour of fluids in the modelled nature. The avoidance of abuses in calibration when using established laws of flow shows us the one side of the story, but this must be complemented by new methods to establish what these laws really are. For example, many hydraulics studies now deal with wetland developments, where a wide variety of kinds of vegetation is encountered, and for calculating the resistance to flow in such cases Manning's n values are often not known, and nor is the adequacy of the Manning formulation at all well established generally. Indeed, as vegetation is deflected to one side by the flow and ultimately flattened, so the resistance must vary in a way that departs markedly from the Manning formulation (Kutija and Hong, 1999 see also Baboric and Keijzer, 1999). Moreover, even in cases where an existing formulation of a resistance law such as that of Manning is adequate, the nature of the bed is not at all well known and may change over time, such as with the season of the year or as a result of the dumping of dredging spill. It is accordingly a first rule of hydroinformatics, corresponding closely to its sociotechnical endeavours generally, that no such change in application practice is possible without the simultaneous provision of new technical means to support this practice. This thesis corresponds to a more general change of paradigm in which a new interpretation and a new light is placed upon 'a set of laws representing all features that are essential and of interest to the problem'.

The data mining paradigm

The formative period of modern science that defined the hydraulics of the twentieth century covered the period between the late fifteenth century and the late eighteenth century. The new foundations were based on the utilisation of the concept of a *physical experiment* and the applications of a *mathematical apparatus* in order to describe these experiments. The works of Brahe, Kepler, Newton, Leibniz, Euler and Lagrange clearly personify such an approach. Prior to these developments, scientific work primarily consisted only of collecting the observables, or recording the '*readings of the book of nature itself'*.

This modern-scientific approach was principally characterised by two stages: a first one in which a set of observations of the physical system were collected, and a second one in which inductive assertion about the behaviour of the system – a hypothesis – was generated. Observational data represent *time- and space-specific knowledge*, whereas a hypothesis represents a *time- and space-generalisation* of this knowledge which *implies* and *characterises* all such observational data. One may argue that, through this process of hypothesis generation, it becomes possible to economise human thought, since more compact ways of describing observations are thereby provided.

As far as water and its movements were concerned, the mathematical formulation of physical phenomena led, along the way, to the theory of ideal fluid motion. But, in parallel, and especially at the beginning of the nineteenth and during the twentieth centuries, the experimental-scientific search for an understanding of nature brought to light the deviations between the behaviours of the observed real world and its proposed conceptualisations. Thus it was impossible to explain such problems as those involving frictional drag or fluid resistance using ideal fluid formulations. Thus, even within this field, two separate domains appeared, theoretical hydrodynamics and hydraulics, with the latter then being defined by Cole in 1962 as "*accumulated semi-empirical data about real fluids in real practical situations, compiled primarily for the engineer*". Most persons involved in hydraulics today would probably consider this definition unacceptable. However it constituted already at that time the dividing line between the engineer, who wished to solve 'real-world' problems, and those mathematicians who held out the prospect of a hydraulics that would die out when all the equations and their solution methods had been developed. Today, at the turning of the twenty first century, we are experiencing yet another change in the scientific process as just outlined. This latest scientific approach is one in which information technology is employed to assist the human analyst in the process of hypothesis generation. This computer-assisted analysis, usually of large, multi-dimensional data sets, is sometimes referred to as a process of *Data Mining for Knowledge Discovery*. The subject of Data Mining for Knowledge Discovery aims at providing tools to facilitate the conversion of data into a number of forms that provide a better understanding of the physical, biological and other processes that generated or produced these data. These new models, when combined with the already available understanding of the physical processes – 'the theory' – result in an improved understanding and novel formulations of physical and other laws, and so provide an improved predictive capability.

As we enter the communicational stage of the digital information era, one of the greatest challenges facing organisations and individuals is how to turn their rapidly expanding data stores into actionable knowledge (Fayed *et al*, 1996). Means for data collection, storage, retrieval and distribution have never been so advanced as they are today. While advances in data storage and retrieval continue at breakneck pace, the same cannot be said about advances in knowledge extraction from large data sets. Without such advances, however, there is a substantial risk of missing what the data has most to offer. The question is thus posed with ever increasing urgency: *what is to be done with all this data?* Ignoring whatever cannot be immediately analysed is wasteful and unwise. This is particularly unacceptable in scientific endeavours, where data usually represents observations carefully collected at considerable expense.

Knowledge Discovery in Databases (KDD) is concerned with extracting useful information from data stores. *Data mining (DM)* is one step (be it fully automated or human-assisted) in this larger KDD process. The broad KDD process includes: retrieving the data from a large data warehouse (or some other source); selecting the appropriate subset with which to work; deciding on the appropriate sampling strategy; selection of target data; dimensionality reduction; cleansing; data mining; model selection (or combination), evaluation and interpretation; and finally the consolidation and the putting to practical use of the extracted 'knowledge'. The data-mining step then fits models to, or extracts patterns from, the pre-processed data.

However, mining the data *alone* is still not the entire story, at least in scientific domains. Scientific theories have long encouraged the acquisition of new data and this data in turn has long led to the generation of new theories. Thus a great deal of 'data mining' has already been done, albeit in a quite other way. When revisiting the earlier methodologies above, we have observed how the traditional process usually began with experimental observations, after which generalisations were postulated, as a theory, and commonly expressed in the form of equations. Thus, traditionally, the emphasis has been on a theory, which demands that appropriate data be obtained through observation or experiment. In such an approach, the discovery process is what we may now refer to as *theory-driven*. Especially when a theory is expressed in mathematical form, *theory-driven discovery* may make extensive use of 'strong' methods associated with mathematics or with the subject matter of the theory itself. The converse view, that is now being so strongly advanced through data mining technologies, takes a body of data as its starting point and searches, using 'weaker' methods, for a set of generalisations, or a theory, to describe the data parsimoniously, and even, possibly but most desirably, to explain it. Usually such a theory takes the form of a precise mathematical statement of the relations existing among the data. This is the *data-driven discovery* process.

We strongly believe that the most appropriate way forward is to combine the best of the two approaches: theory-driven, understanding-rich, processes with data-driven discovery processes.

Model Induction

Histories of science, and especially those of mathematics, draw a particular attention to the development of a physical symbol systems, such as a scheme of notation in mathematics, interactively with the evolution of more refined representations of physical and conceptual processes in the form of equations in the corresponding symbols. It is then a common experience that one and the same physical symbol system, may serve for the expression of a great number of different equations. To the extent that each equation can be regarded not only as a string of symbols that can be manipulated mathematically, but also interpreted as a collection of signs which serves as a sign that points towards a particular physical object, process or event, so it constitutes a *model* of that object, process or event (Abbott, 1992; 1993; see also Klinkenberg, 1996, p. 180). It is then usual to speak of a collection of 'indicative' signs that, through this collectivity, points towards an 'expressive' sign. Data, on the other hand, remain as 'mere' data just to the extent that the set of data constitutes a collection of indicative signs that does not serve as an expressive sign, so that it does not point immediately to anything meaningful. From this point of view, the evolution of an equation within a physical symbol system as a means of better conveying the 'meaning' or 'semantic content' that is encapsulated in the data corresponds to the evolution of another kind of sign which does express something to us, and thereby defines a model. Evidently the 'information content' is very little changed, or even unchanged, when a body of data is transformed into an equation derived from this data, but the 'expressivity' or 'meaning value' is commonly increased immensely. Since it is just this increase in 'meaning value' that justifies the whole activity of substituting equations for data, there is a natural interest in processes for further promoting such means for effecting what are again essentially 'economies of thought'.

Model induction is one particular mode of data mining. Inferring models from data is an activity of deducing a closed-form expression based solely on observations. Observations, however, always represent (and in principle only represent) a *limited source of in-*

formation. The question then has to be posed of how the corresponding limited flow of information from a physical system to an observer can result in the formation of a model that is complete in the sense that it can account for the *entire* range of phenomena encountered within the physical system in question – and so a model that can describe even the data that are outside the range of previously encountered observations (Babovic, 1996b). Now traditional model induction can often be usefully characterised as the search for a model that is capable of *acquiring semantics from syntax*. Clearly, every model has its own syntax. Artificial neural networks, for example, have the syntax of a network of interconnected neurons, whereas genetic programming has the syntax of tree-like networks of symbolic expressions in reverse Polish notation, or RPN. The question is whether a particular syntax can capture the semantics of the system that it attempts to model. Certain classes of model syntax will surely be inappropriate to the representation of some physical systems. One might try to choose the model whose representation is as complete as possible, in the sense that a sufficiently large model can capture the data's properties to a degree of error that decreases with an increase in the model size. Thus, to revert to the standard methods of the earlier practice, one might decide to expand in Taylor or Fourier series to a degree that will decrease the least-squares error to a certain, arbitrarily given degree. However, in most of the cases of interest here the semantics would almost certainly not be captured using the syntax of such methods (see, more generally, Klinkenberg, 1996, pp 143-154).

Genetic Programming

Genetic Programming (Koza, 1992) is one instance of the evolutionary algorithms family. In Genetic Programming (GP) the evolutionary force is directed towards the creation of representations, often called 'models' that take a symbolic form. In fact the strings of tokens, such as constitute equations for example, although usually treated as strings of symbols and employed as such, can also be regarded as sequences of sign in the strict sense. In this last situation they can be regarded as models whose expressive signs provide definite meanings. Thus, although the term 'symbolic model' is an oxymoron in the strict sense of the term of semiotics – and indeed even more so than the term 'data model' – it can still be employed with this special understanding. In this evolutionary paradigm, evolving entities are presented with a collection of data and the evolutionary process is directed towards the creation of such closed-form symbolic expressions describing the data. In its primitive form, GP lends itself quite naturally to the process of induction of mathematical models based on observations: GP is an efficient search algorithm that need not assume the functional form of the underlying relationship. Given an appropriate set of basic functions, GP discovers a (sometimes very surprising) mathematical model that approximates the data well. At the same time, GP-induced models come in a symbolic form that is familiar to many scientists so that their production can be more readily assimilated as 'knowledge' as soon as their 'symbols' are in fact regarded as signs. (see, for example, Babovic, 1995).

GP iteratively applies variation and selection on a population of evolving tree structures standing for symbolic expressions in RPN. Standard variation operators in genetic programming are subtree mutation (replace a randomly chosen subtree with a randomly generated subtree) and subtree crossover (replace a randomly chosen subtree from a formula with a randomly chosen subtree from another formula). For a detailed description, see, for example, Babovic and Abbott (1997a). The types of functions used in this tree structure are user-defined. This means that they can be algebraic operators, such as *sin, log,* +, -, etc., but they can also take the form of *if-then-else* rules, making use of logical operators such as *OR, AND,* etc. A number of applications of GP has been reported, such as studies by Babovic and Minns (1994) in which salt intrusion data were analysed, Babovic (1995) related to experimental data for bed concentration of suspended sediment, and Babovic (1997) related to rainfall runoff modelling. In all of the above-mentioned studies, GP-induced relationships provided more accurate descriptions of data than those obtained using more conventional methodologies. An extensive survey of the applications of GP in water resources is provided in Babovic and Abbott (1997b).while recent issues of the *Journal of Hydroinformatics* provide other examples and references.

However, the application of standard GP in a process of scientific discovery does not always guarantee satisfactory results. In certain cases, GP-induced relationships are too complicated and provide little new physical insight into the process that generated the data. One may argue that GP, in such situations, blindly fits parse trees to the data (in almost the same way as in Taylor or Fourier series expansions). It can be argued that GP then results in a model with accurate syntax, but with opaque, or even meaningless, semantics. Moreover, in these cases, the dimensions of the induced formulae often do not fit, pointing to the physical inadequacy of the induced relationships and the presence of further variables 'hidden' within the coefficients.

Dimensionally Aware Genetic Programming

Recently, an augmented version of GP has been proposed – dimensionally aware GP – which is arguably more useful in the process of scientific discovery (Keijzer and Babovic, 1999; Babovic and Keijzer, 1999).

In all scientific endeavours data representing carefully collected observations about particular phenomena that are under study are usually accompanied by their units of measurement. However, the traditional methods usually exploit this information only through the introduction of dimensionless ratios (with such well known examples as Froude and Reynolds numbers). Once the dimensionless numbers are used instead of the original dimensional values, the problem of dimensional homogeneity is conveniently avoided, as all analysed quantities are dimension-free. It can also be argued that dimensionless ratios reduce the dimensionality of the original search space, making it more compact and thus providing a more effective behaviour of algorithms that fit models to the data. At the same time, however, much of the information originally contained in the units of measurement of the data sets is ignored entirely.

Standard GP is ignorant of the dimensionality of its terminals and can safely be applied to problems composed of dimensionless numbers only. Given the symbolic nature of GP and its ability to manipulate the structure of functional relationships, it seems strange that information contained in units of measurement has not been used earlier as an aid in the search process. After all, the criterion of dimensional correctness as used in science acts as a syntactic constraint on any formula it induces. It was therefore to be expected that the introduction of dimensions into the GP paradigm ought to result in an improved search efficiency. The example of the Manning roughness coefficient *n* may again be used.

Now it appears quite generally that the dimensions of some of the best known parameters used in hydraulics are anything but well understood and 'transparent'. In the case of the example of the Manning number, this has dimensions of $s^1\ m^{-1/3}$ but is still referred to as a roughness 'coefficient'! The fact is that Manning's *n* must accommodate a number of phenomena not explicitly taken into account in the well-known formula defining average flow velocity under steady conditions.:

$$u = \frac{1}{n} R^{2/3} I^{0.5} \tag{1}$$

Since functional similarity to the natural system is supposed to be comprehended by equation (1) itself, it is this *calibration coefficient n* that must capture the exact correspondence between the model and the real world. As explained above, parameters of this sort serve in effect as *error compensation devices* that artificially adjust the model results to compensate for the fundamental discrepancies that exist between the real world and its representation within the model. Manning's *n* is much more than a coefficient that can be associated with roughness-induced forces only, and as such is not well defined in the natural world. Instead it exists only at the interface between nature and model. It has to accommodate both: a part related to physical processes, but also a part that has to do with our schematisation of nature within a model. One may even ask 'What is the physical meaning of such a parameter – how well is it grounded, and indeed is it properly grounded at all?' We may be able to read a certain 'physical meaning' into such calibration parameters, but they do not exist-as-such and are thus 'disconnected' in a fundamental way from the world that they are supposed to model (Minns and Babovic, 1996). At the same time, it is quite obvious that the dimensions of such calibration coefficients will then be chosen in such a way that, although the overall dimensional consistency of the model will be maintained, the coefficients themselves may have little physical meaning.

Thus, within the data -mining paradigm it is possible to induce a completely new definition of coefficients, such as roughness. Such new definitions must then however be supported not only by collected observations but also by physical insights. In this way a descriptive, semantic component is added to the data-mining algorithm. This is in addition to the functional semantics that define the manipulations on numbers. While functional semantics grounds formulae in mathematics, the dimensional semantics grounds them in the physical domain. Hopefully, the kinds of misunderstandings outlined in the case of roughness at the beginning of this paper can then be prevented.

It should be emphasised here that in the case of data mining for knowledge discovery in particular, the hydroinformatician must have *both* a sound insight into the functioning of the tools that are employed *and* a profound insight into the physical processes towards which the data mining is directed. Whereas in the symbolic era the most difficult part of the work was often at its beginning, in the formulation of the basic equations, in this new paradigm the greatest difficulties usually arise at the end of the process, in interpreting and 'making sense' of the productions of the data mining tools, so that these productions come to constitute true 'knowledge'. It is then necessary also that the hydroinformatician can communicate meaningfully with experts in the domain of application so that this new knowledge can be related to existing knowledge and specifically to existing theory.

Further Influences of Hydroinformatics

We have traced the beginning of the new, hydroinformatics, paradigm in hydraulics itself to two streams of development, the one occurring outside and the other within hydraulics. The first of these, the one that is imported from outside of hydraulics, is that of data mining for knowledge discovery. The second, which has arisen within hydraulics, is that of the accumulation of massive amounts of data, often collected at great expense, which data is eminently suited to data mining for knowledge discovery. However, since the hydroinformatics paradigm is also one of combining technologies in order to lever added value from out of their combination, calibration procedures are nowadays increasingly supported by field and laboratory measuring programmes, together with new technologies for mining the resulting field data specifically for calibration purposes. These practices are commonly complemented by the direct assimilation of data into numerical models (e.g. Canazares, 1998; see, more generally, Abbott, 1996). For example, acoustic-doppler records can now be mined to determine the nature of the bed material and forms, thus supporting the modelling process in a more satisfactory way. Insofar as unsupported calibration proceeds at all, it is now increasingly in its turn automated, so as to be so much less 'lumped' - or so much more 'distributed'- thereby highlighting areas in which the calibration procedure itself becomes unrealistic. Thus, the process of discovering new knowledge becomes 'data driven' in other ways again. It is in the combination and integration of such ways that the process of creating knowledge in hydraulics is expected to proceed for the greater part in 'the hydraulics of the hydroinformatics era'.

The sociotechnical dimension of hydraulics

As a technology, hydraulics has always been associated with activities in the 'outer', or 'material' world of applications with a social utility. For example, in 1786 the French military engineer, Pierre Louis Georges du Buat, published the second edition (the 1779 first edition was not complete) of his major work, *Principes d'Hydraulique*. He enumerated there what he considered to be the most urgent list of problems which hydraulics should solve:

> "Our understanding of hydraulics has been extremely limited:.. We are.. in almost absolute ignorance of the true laws to which the movement of water is subject: .... All that concerns the uniform course of the waters of the surface of the earth is unknown to us; and to obtain an idea of how little we do know, it will suffice to cast a glance over what we do not.
>
> "To estimate the velocity of a river of which one knows the width, the depth and the slope; to determine to what height it will rise if it receives another river in its bed;.. how much it will fall if one diverts water from it; to establish ... the proper capacity of the bed to deliver to a city at a given slope the quantity of water which will satisfy its needs; to lay out the contours of a river in such a manner that it will not work to change the bed in which one has confined it; to calculate the yield of a pipe of which the length, the diameter, and the head are given; to determine how much a bridge, a dam or a gate will raise the level of a river; to indicate to what distance backwater will be appreciable, and to foretell whether the country will be subject to inundation; to calculate the length of a canal to drain marshes long lost to agriculture; to assign the more effective form to the entrances of canals, and to the confluences or mouths of rivers; to determine the most advantageous shape to give to boats or ships ..; ... All these questions, and infinitely many others of the same sort, are still unsolvable; who would believe it?
>
> ".. for lack of principles, one adopts projects of which the cost is only too real but of which the success is ephemeral; one carries out projects for which the goal is not attained; one charges the state, the provinces, the communities with considerable costs, without gain, often with loss; or at least there is no proportion between the cost and the advantages which results therefrom". [Du Buat, 1786; quoted from Rouse and Ince, 1957]".

Not only are the meaning and language intelligible to any literate citizen but they also correspond to social demands. Moreover, the domain of hydraulics defined on this basis of social application is already very wide indeed. It was just the century which followed Du Buat's observations, however, which witnessed the explosion of the mathematical sciences and, in particular, all the various attempts to formulate real-world physical phenomena in terms of equations which might in principle be solved and thus might open the way to engineering applications - with all the restrictions that this approach also entailed.

It was essentially this mathematical-scientific development of hydraulics that led to the gradual separation of the hydraulician from his or her immediate social environment: it thereby became possible for many hydraulicians to concentrate their attention upon the one or the other technical, and even scientific, aspect of a problem, while leaving the social aspects and even the constructional aspects to 'other people' (Abbott, 1999). The hydraulician could retreat into his or her laboratory and office and follow the principle that 'what is good for hydraulics and flow efficiency must also be good for the mankind and nature'. Oscillating unconsciously between the 'beaver mentality' commonly attributed to the engineer and the arrogance of an ancient Egyptian priest, the hydraulician increasingly usurped the decision power over water resources in society while becoming at the same time increasingly detached from the social and ecological consequences. In effect, although it was supposed that the technical results of studies would somehow and at some time find social applications, the processes whereby such social applications occurred were not seen as the direct concern of the hydraulician and these processes were even less themselves regarded as subjects of hydraulics research. In the new paradigm, on the other hand, *it is practically impossible to research in hydraulics at all without simultaneously researching how the results of this research will enter into social applications*. Within this new paradigm, new technological means can scarcely be developed to any effect at all without a thorough research being made into the ways in which these new means can best be applied within society. Entirely consentaneously, no changes can be realised and sustained in society without the elaboration of new technical means to catalyse and guide these changes. Hydroinformatics is an essentially *sociotechnical* endeavour, and hydraulics is carried along with hydroinformatics in this endeavour (Thorkilsen and Dynesen, 2001).

Thus the era when the hydraulician could concern his or herself exclusively with the technical aspects of the project, leaving its social application aspects to 'other people' has largely come to an end. In the language of sociotechnical studies, the hydraulician is no longer a homogeneous engineer, but a heterogeneous one. Of course there have always been heterogeneous engineers in practice (Edison and Bell commonly serve as exemplars on the sociotechnical literature) but in earlier times these could be regarded as exceptions. Nowadays, heterogeneity is the rule (Abbott, 1996).

In the same vein, the hydraulician, in his or her role as a hydraulic engineer, might previously have travelled to a project site to establish 'the facts'. This person was not normally called upon to enter into the making of judgements about the social desirability of the project to which the facts applied and even less to make arrangements for such judgements to be expressed by local persons, and even less again to research, analyse, design and install the integrated social and technical, or sociotechnical, arrangements most appropriate for this purpose. Today, the hydraulician, as a hydraulic engineer working within the hydroinformatics paradigm, is in the first place one who *persuades* the rest of society to follow one line of action rather than others, whether individually, or (as is much more common nowadays) as a team with this same purpose, backed by an veritable arsenal of technical equipment. And, increasingly, this line of action involves the indivisible provision of the technical and the social, or sociotechnical, means for the ordinary, and most commonly non-technical, persons directly concerned to make this choice of the one line of action rather than other such lines by interacting with one another in further persuading activities, In effect, the task of the hydroinformatician becomes increasingly one of 'persuading people to persuade people' (Abbott, 1999 a and b; Abbott and Ionoski, 2001).

The decisive role of knowledge management

The most important single factor by far in the current development of hydroinformatics is the exploitation of the potential of the Internet and such electronic networks, or 'internets' with a lower case 'i', that are coming to succeed it. Indeed, it is ultimately through its application of internet-based technologies that hydroinformatics comes to be legitimised as an independent discipline. In terms of the post-modern distinction between 'knowers' and 'consumers of knowledge', the most immediate effect of such networks is to increase the number of consumers of the most advanced hydraulic knowledge by at least two orders of magnitude again. The development in the number of such users and its future projection is also followed in Fig. 1. However, sociotechnical studies show again that such a massive change or 'paradigm shift' in the social application of hydraulic and related knowledge cannot be realised without the introduction of new technical means. In particular, insofar as non-expert members of the general public wish to make use of such knowledge in order to assess the impact of various projects (ground water mining, river embankment construction, urbanisation etc.) upon their quality of life, so they need new classes of tools to assist them. They need tools that will enable them to relate their own personal and group interests, concerns and intentions to any proposed intervention in their environment if they are to participate at all effectively in the judgmental processes governing such interventions. Any tool that enables persons and groups, whether expert or non-expert, to make judgements on the basis of presented facts is called a *judgement engine* (Abbott, 1998b; Abbott, 1999; Huang at al, 1999). Thus a judgement engine is a computational device that facilitates the making of judgements in a decision-making environment. The decision making process is commonly supposed to entail strings of inferences, typically of the form:

$$\begin{aligned}&\textit{(beliefs, facts (data))} \rightarrow \textit{attitudes} \rightarrow \textit{positions} \rightarrow \textit{judgements}\\ &\rightarrow \textit{decisions} \rightarrow \textit{actions}\end{aligned} \tag{2}$$

Here the inferential processes are constructed as mappings, denoted by arrows, from one specific state of conscious activity to another. Each such state is specified by its type, as described by a noun in a natural language in (2) above. As its own name suggests, a judgement engine is one that facilitates the process described in (2) reading up to the level of judgements from left-to-right, so that it constitutes one (of many) representations of a judgmental process.

In so far as a number of persons participate in a judgmental process, so a great variety of beliefs, attitudes, positions, etc will be presented, but there is still an interest in maintaining a community of structure in the overall processes exemplified in (2). This notion of communality of structure can be made more explicit by associating it with the requirement that all inferential strings such as (2), when taken over all persons participating in the inferential process, should constitute a category in a specific mathematical sense (Abbott & Dibike, 1998a and b; see also Abbott, 2000b, where the judgement engines must be instantiated automatically from 'user profiles' which characterise the requirements for advice of each individual user.).

The problem of promoting cooperation between persons participating in an inferential process can then be posed as one of inducing an understanding of the origins of the judgements of other participants. This implies, to take the example (2), that the category that is the dual of (2),

$$\begin{aligned}&\textit{actions} \rightarrow \textit{decisions} \rightarrow \textit{judgements} \rightarrow \textit{positions} \rightarrow \textit{attitudes}\\ &\rightarrow \textit{(beliefs, facts (data))}\end{aligned} \tag{3}$$

should be made as explicit as possible when 'actions' and 'facts(data)' are observables. A judgement engine that facilitates this process is said to be *transparent*.

It should be observed that judgmental processes occur through the process described in (2) and (3) through the presence of many other processes than those represented explicitly by the mapping themselves, but that these belong to a different class of judgements - which we may call 'intermediate judgements' - than those identified explicitly in these strings of inferences (See, again Abbott, 1999, and Huang *et al* 1999; and then, originally, Husserl 1938/1938//1957)). Since such engines are designed explicitly for use over electronic networks, they can best be appreciated through their hands-on use over an electronic network. (See, again, Huang et al, 1999)

Conclusions

Hydroinformatics cannot be regard merely as a new way of applying an otherwise unchanged hydraulics within society. Instead, hydroinformatics, while building upon existing hydraulics and taking existing hydraulics for the most part as its central core, introduces its own paradigm into hydraulics itself. Some of the consequences of this introduction have been elaborated here, but there are increasingly many others. Hydroinformatics takes up many activities of hydraulics research and practice that were hitherto considered as rather separated and brings them together in new and innovative ways using the new information and communication technologies. By these means it increases the value of the individual activities, often markedly. Thus, for example, the bringing together and integration of field measurements, numerical models and remotely sensed data provide new and highly valuable new facilities that increase greatly the value of the component activities. Indeed, studies of such biological-geographical problems as those of following changing distributions of eel grass plains and mussel beds under the impacts of depositions of fine sediments are scarcely possible without such integrations (Abbott, 1997; Thorkilsen and Dynesen, 2001). Hydroinformatics is strongly engaged in these combining and integrating activities. Such procedures then lead to new ways of doing hydraulics itself, as exemplified in this paper.

Further to this again, hydroinformatics also introduces new techniques to encapsulate existing knowledge, commonly with the view to accelerating the rate of access to this knowledge. Artificial neural networks are already widely used for this purpose (e.g. Dibike et al., 1999). Such developments drive several new areas of applications of existing hydraulics knowledge, such as real-time control of urban drainage and irrigation systems, real-time diagnostic systems and on-line risk analysis support systems. Just as the information revolution has already led to profound changes in the ways of working of hydraulicians, so the ongoing communication revolution is changing the whole relation between the hydraulician and society, From being one who personally 'thinks about water', the hydraulician is drawn (or even dragged!) by degrees to become one who has also to 'think about how other persons think about water''. Although the more purely technical developments in hydroinformatics, as exemplified here by data mining for knowledge discovery, are the most readily acceptable to the more traditionally minded hydraulician, the sociotechnical developments promise ultimately to lead to the most profound changes in the practice of hydraulics itself.

References

Abbott, M. B., 1991, *Hydroinformatics: information technology and the aquatic environment*, Ashgate, Aldershot, UK, and Brookfield, USA.

Abbott, M. B., 1992, The theory of the hydrologic model, or: The struggle for the soul of hydrology, in *Advances in theoretical hydrology; a tribute to Jim Dooge* (ed. O'Kane, J. P.) pp 237-254, Elsevier, Amsterdam.

Abbott, M. B., 1993, The electronic encapsulation of knowledge in hydraulics, hydrology and water resources, *Adv. Wat. Resources*, **16**, pp 22-39.

Abbott, M. B., 1994, Hydroinformatics: a Copernican revolution in hydraulics, *J. Hyd. Res., Special Issue on Hydroinformatics*, pp 3-13.

Abbott, M. B., 1996, The sociotechnical dimension of hydroinformatics, in *Hydroinformatics '96* (ed. Mueller, A.) pp3-18, Balkema, Rotterdam.

Abbott, M. B., 1997, Range of tidal flow modelling, *Journal of Hydraulic Engineering*, ASCE, Vol 123 (4): pp. 257-277.

Abbott, M. B., 1998a, New business opportunities; new human responsibilities, in *Hydroinformatics '98* (eds. Babovic, V. M. and Larsen, L.-C.) pp 1171-1175, Balkema, Rotterdam.

Abbott, M. B., 1998b, Hydraulik im Zeitalter der Hydroinformatik, in Steinebach, G., *Mathematische Modelle in der Gewaesserkunde*, BFG, Koblenz, pp 126-129.

Abbott, M. B., 1999a, Introducing Hydroinformatics, *J. of Hydroinformatics*, **1**, 1, pp3-19.

Abbott, M. B., 1999b, Forchheimer and Schoklitsch: a postmodern retrospection. In memorial symposium P. Forchheimer and A. Schoklitsch, 1ed. Graff, W. H., Proc. 28th IAHR Congress, Graz.

Abbott, M. B., 2000a, Fides et Ratio, a critical appraisal (Book review: Faith and Reason: Evangelical letter *Fides et Ratio* of the supreme Pontiff John Paul II etc.) *Journal of Hydroinformatics* , **2**, 2, in the electronic version only, but with printed proofs available from the author.

Abbott, M. B., 2000b, The gender issue in hydroinformatics, *Journal of Hydroinformatics*, **2**, 2, pp 87-106.

Abbott, M. B. and Ionoski, A., 2001, The democratisation od decision making processes in the water sector, *J. of Hydroinformatics*, 3 (1), pp 35-48.

Abbott, M. B., and Dibike, Y, 1998a, On the representation of processes in the postsymbolic era, in *Hydroinformatics '98* (eds.Babovic, V. M., and Larsen, L. C.) pp1177-1184, Balkema, Rotterdam.

Abbott, M. B., and Dibike, Y, B.,1998b, The symbolic representation of hydroinformatics processes using elements of category theory, in *Hydroinformatics '98* (eds. Babovic, V. M. and Larsen, L. C.) pp 1185-1192, Balkema, Rotterdam.

Abbott, M. B. and Jonoski, A..,1998, Promoting collaborative decision making through using electronic networking, *Hydroinformatics 98*, ed. Babovic, V., and Larsen, L. C., pp911-918, Balkema, Rotterdam.

Appeganesi, R., and Garratt, C, 1995, *Postmodernism for beginners*, Icon, Cambridge.

Babovic, V., 1996a, *Emergence, Evolution, Intelligence; Hydroinformatics*, Balkema, Rotterdam

Babovic, V., 1996b., "Can water resources benefit from artificial intelligence?", in Köngeter,J., (ed.), *Internationales Wasserbau Symposium – Computational Fluid Dynamics; Bunte Bilde in der Praxis*, Aachen, Germany

Babovic, V., 1997, "On the modelling and forecasting of nonlinear systems", in Refsgård, J.C., and Karalis, E.A., (eds), *Operational Water Management*, Balkema Publishers, Rotterdam

Babovic, V., and Abbott, M.B., 1997a., "The evolution of equations from hydraulic data, Part I: Theory", *Journal of Hydraulic Research*, **35,** 3, pp. 1-14

Babovic, V., and Abbott, M.B., 1997b.,"The evolution of equations from hydraulic data, Part II: Applications", *Journal of Hydraulic Research*, **35,** 3, pp. 15-34

Babovic, V., Larsen, L.C. and Wu, Z**,** 1994., "Calibrating Hydrodynamic Models by Means of Simulated Evolution" in *Hydroinformatics '94*, Verwey, A., Minns, A.W., Babovic, V., and Maksimovic, C., (eds.), pp.193-200.

Babovic, V., and Keijzer, M., 1999, Genetic Programming as a Model Induction Engine, Journal of Hydroinformatics, **2,** 1, pp 35-60.

Babovic, V., and Minns, A.W., 1994, "Use of computational adaptive methodologies in hydroinformatics", in Verwey, A., Minns, A.W., Babovic, V., and Maksimovic, C., (eds.), *Proceedings of the 1st International Conference on Hydroinformatics, Hydroinformatics '94,* pp. 201-210

Baudrillard, J., 1963, *Le systeme des objets: la consommation des signes*, Gallimard, Paris.

Dibike, Y. B., Minns, A. W., and Abbott, M. B.,1999, Applications of artificial neural networks to the generation of wave equations from hydraulic data, *J. Hydraulic Res.* **37**, 1, pp81-97.

Eco, U., 1976, *A theory of semiotics*, Indiana Univ., Bloomington, USA.

Huang, Y., Solomatine, D., Velickov, S., and Abbott, M. B., Distributed environmental impact assessment using internet, *J. of Hydroinformatics*, **1**, 1, pp59-70.

Husserl, E, 1938//1957, *Experience and judgement; investigations in a geneology of logic* (trans. Churchill, J. S., and

Ameriks, K.) Routledge, London.

Jonoski, A, and Abbott, M.B., 1998, Network distributed decision support systems as multi agent constructs, in *Hydroinformatics '98* (eds. Babovic, V. M., and Larsen, L.-C.) pp 1219-1226, Balkema, Rotterdam.

Keijzer, M., and Babovic, V., 1999, Dimensionally aware genetic programming, in Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Honavar, V., Jakiela, M., & Smith, R. E. (eds.)., GECCO-99: *Genetic and Evolutionary Computation Conference*, Morgan Kaufmann, Orlando and San Francisco.

Klinkenberg, J.-M., *Précis de sémiotique générale*, De Boeck et Larcier, Paris.

Koza, J., 1992., *Genetic programming: on the programming of computers by natural selection*, MIT Press, Cambridge, Mass.

Kutija, V., and Hong, 1996, A numerical model for assessing the additional resistance to flow induced by flexible vegetation, *J. Hyd. Res.*, 34 (1), pp 99-114

Lyotard, J-F., 1979//1992, *The postmodern condition, a report on knowledge* (trans. Bennington, G., and Massumi, B.) Manchester Univ., UK.

Lyotard, J-F, 1998, *Le postmoderne explique aus enfants, correspondence 1982-1985,* Galilee, Paris.

Piatetsky-Shapiro, G., and Smyth, P., 1996., From data mining to knowledge discovery: an overview, in *Advances in Knowledge Discovery and Data Mining*, Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., eds., MIT Press, Cambridge, Mass., pp.1-36.

Rouse, H. and Ince S., 1957, *History of Hydraulics*, Dover, New York.

Thorkilsen, M., and Dynesen, C., 2001, An owner's view of hydroinformatics: its rule in realising the bridge and tunnel connection between Denmark and Sweden, *J. Hydroinformatics*, 3 (2), pp 105-135.