A model selection protocol to support the use of models for water management

D. B. Boorman, R. J. Williams, M. G. Hutchins, E. Penning, S. Groot, J. Icke

▶ **To cite this version:**

D. B. Boorman, R. J. Williams, M. G. Hutchins, E. Penning, S. Groot, et al.. A model selection protocol to support the use of models for water management. Hydrology and Earth System Sciences Discussions, 2007, 11 (1), pp.634-646. hal-00305646

HAL Id: hal-00305646
https://hal.science/hal-00305646

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hydrol. Earth Syst. Sci., 11(1), 634–646, 2007
www.hydrol-earth-syst-sci.net/11/634/2007
© Author(s) 2007. This work is licensed
under a Creative Commons License.

Hydrology & Earth System Sciences

A model selection protocol to support the use of models for water management

David Boorman[1], Richard Williams[1], Michael Hutchins[1], Ellis Penning[2], Simon Groot[2] and Joost Icke[2]

[1]Centre for Hydrology and Ecology, Wallingford, Oxfordshire, OX10 8BB, UK
[2]WL | Delft Hydraulics, P.O Box 177, 2600 MH, Delft, The Netherlands

Email for corresponding author: dbb@ceh.ac.uk

Abstract

Models are widely used to assist water managers in their duties, yet there is little advice to help water managers select a model appropriate to their particular application. The advice that is available is specific to either the application type or models considered, focused on one aspect of model application, e.g. goodness-of-fit comparisons between observations and simulations and unrealistic, e.g. in separating model selection from wider aspects of the intended application. This paper proposes a model selection protocol that recognises the particular requirements of the application, the roles of individuals associated with modelling studies, and the importance of establishing an informed dialogue between them. Establishing this common understanding at the outset of any modelling study is highly important in securing a successful outcome to the modelling process.

Keywords: modelling, water management, model selection, Water Framework Directive

Introduction

Models are now recognised by many water managers as tools that can be of great assistance in the performance of their duties. Models can provide: an insight into the way rivers and catchments function; a framework for data analysis; a means of investigating problems; and, perhaps most usefully in the context of this paper, a way of evaluating alternative management strategies (Van der Molen, 1999; Wasson *et al.,* 2003; Moll and Radach, 2003). However, using a model to support a water management function is not straightforward. A water manager must identify which model to use, who should use it, know how to interpret the outputs, and how to present the results to a wider stakeholder community.

At present, useful guidance exists that helps with the process of model application, e.g. Good Modelling Practice (Van Waveren *et al.*, 1999), and this is being further developed within an on-going research project part-funded under the European Commission 5th Framework Programme, HarmoniQuA (e.g. Refsgaard and Henriksen, 2004; Refsgaard *et al*., 2005). This guidance takes the form of a step-by-step approach to the modelling process so that at any stage it should be clear what has been achieved and what to do next. Such guidance is relevant once the modelling process is underway, but at the start of this process it is necessary to select the model to be used. A review of published information reveals that while much information is available describing models and their application, there is little practical advice on model selection.

This paper presents a model selection protocol to support the use of models for water management. The protocol is not a scoring system to identify the best model; instead, the objective is to achieve the best outcome from the model application. It is proposed that this is achieved most effectively by properly specifying the characteristics and requirements of the study and then comparing these with what the model can deliver. This technical and objective process depends, inevitably, on the skills, expertise, experience and expectations of those involved. The protocol provides a framework for a broad discussion that will minimise confusion and misunderstanding and establish an effective foundation for the modelling study.

Background

A great many publications describe the application of a single model where they have, among other things, presented what has been achieved with a modelling approach in a particular circumstance (Bantz and Cembrowicz, 1983), introduced novel ideas in model application, e.g. parameter calibration (Van Oijen *et al*., 2005), or revealed the greatest uncertainties in determining management options (Warwick and Roberts, 1992). The reader can infer that because a model or technique has been used successfully in the application reported, it may have more general applicability but such papers are not intended to enable a comparison between models that will help in model selection.

However, many publications do report on model comparisons and, generally, these focus on a relative assessment of goodness-of-fit indices. The studies range from the straightforward, such as Crabtree *et al.*, 1986 (two models compared on one study site) and Hughes, 1994 (four models on one site), to the complex, Houghton-Carr, 1999 (four models on 25 study sites) and Vanderwiele and Ni-Lar-Win, 1998 (11 models on 55 study sites). The more complex comparisons generally conclude that there is no single best model. The more limited comparisons deliver more specific differentiation between models and often do recommend one model over the others. However, there are often caveats; for example, in both the studies referenced above, readers are advised to note the importance of input data as much as differences in model performance, e.g. "the major difficulty in applying both models is the availability of sufficient …data" (Crabtree *et al*., 1986).

Other publications offer more generic advice on how models should be compared. Bart (1995) suggests this should be based on descriptions of the models and estimates of the reliability of the predictions. Green and Stephenson (1986) advise on goodness-of-fit criteria and conclude that the most appropriate criterion will depend on the purpose of the application. Van der Perk (1997) contains a method for comparing model structures based on the Akaike information criteria, and Thian Yew Gan *et al*. (1997) suggest that more emphasis should be given to structure and data issues than to a consideration of model complexity. Together these publications contain many of the elements that need to be included when comparing models with a view to selecting one for a particular application.

Useful advice on the process of model selection can be found in the 'grey' literature. For example, the US EPA (1987) places model selection in the context of the whole modelling process, from problem definition to model application. The model selection criteria are in three parts: objectives criteria, technical criteria (i.e. relating to model processes) and implementation criteria (i.e. model quality assurance and documentation). Further guidance from the EPA in the context of total maximum daily load modelling is presented by Shoemaker *et al*. (1997); this goes beyond technical aspect of modelling to consider issues such as the availability of trained personnel, the long-term commitment to the model and in-house modelling experience. The basis of this advice is attributed to a chapter, "Mathematical modelling of the combined sewer system", in a book *Control and treatment of combined sewer overflow* (Nix, 1990), not an obvious source for guidance on model selection. The internet now provides another source of 'grey' literature and the information available in the *Series on Model Choice* at www.toolkit.net.au/modelchoice provides much practical advice under four headings: objectives, data, expertise and resources. Particularly interesting are the comments in the section: *So what tends to happen in reality?* "…commonly, the objectives are not entirely clear and one uses gut feeling, or chooses whatever model is most easily available (or have used before) or perhaps even whatever might most impress the client!".

In a recent publication in this area, Saloranta *et al.* (2003) claim that their model selection criteria "will provide the growing number of (non-specialist) model users the means by which to judge from the model user's perspective … the appropriateness of a given model for the particular water management application". While many of the ideas presented by Saloranta *et al*. are helpful, their criteria fail to recognise either the roles and responsibilities of those involved in modelling studies, or the widely different characteristics of modelling studies.

An overview of roles and models

The requirement for a modelling study usually comes from the person responsible for the management or protection of the water resource. It may be that the requirement is in fact driven by others, for example those wishing to exploit, or protect, the resource. The frequently used term 'end-user' blurs these two roles. In this paper, the term **water manager** will be used to describe the person responsible for the study and its outcome; this person might be considered to 'own the problem'. The other interested parties, i.e. those with interests in exploiting or protecting the resource, will be described as **stakeholders**. It is likely that the water manager is also a stakeholder, but the important point here is the special role that the water manager has in delivering the success of the study. The water manager therefore specifies the problem, or research issue, that needs to be addressed to meet the requirements of all the stakeholders.

The modelling activity would normally be undertaken by

a specialist with expert knowledge of both the issue, in a general sense, and valid modelling approaches; it is obvious to refer to this person as the **modeller**. The modeller must apply the model, present results and interpret the model outputs. The water manager will then decide on the appropriate management action based on information and results supplied by the modeller and, in many cases, other expert advice.

Note that in the above description, and throughout this paper, the three roles, i.e. stakeholder, water manager and modeller, have been assigned to individuals, but it would be equally valid to see these as functional roles of an organisation, or a department within an organisation.

So far in the discussion, it has been convenient to use the term 'model' in a number of different ways. It has been used to refer to the model system, usually computer software, and the use of the model for a specific study. Refsgaard and Henriksen (2004) and Van Waveren *et al.* (1999) argue for a clear terminology when discussing modelling issues. They suggest distinguishing between the model code and the model application. The **model code** is "a mathematical formulation in the form of a computer program that is so generic that it, without program changes, can be used to establish a model with the same basic type of equations (but allowing different input variables and parameter values) for different study areas". A model application is "a site-specific model established for a particular study area, including input data and parameter values". Thus the combination of a model code with a particular dataset and, if necessary, calibration, is called a **model application.**

It is possible to make further distinctions within the broad term model, for example to recognise the use of particular **model algorithms** to solve the **model equations**. However, for the purpose of this paper it is sufficient just to distinguish between **model codes** and **model applications**.

What is the process by which a **water manager** commissions a **model application**? Saloranta *et al.* (2003) suggest that using their criteria a non-specialist can select the model code, and that a code that requires a specialist user may be inadequate, implying that the non-specialist will also take on the role of modeller. The *Series on Model Choice* (2005) suggests the modeller chooses the model and then justifies the selection to the client. The authors of this paper agree that the water manager is more likely either to choose a modeller, or a combination of modeller and model code, than choose the model code itself. In doing so, the manager recognises the expertise of the modeller, not only in developing the model application but also in selecting, or justifying the selection of, a model code for a particular case study. However, most modellers will have expertise with only a sub-set of the available model codes. Thus, when asked to select a model code for a particular case study, the first ones the modeller will consider are likely to be those with which he or she is most familiar. Only when these are all rejected will a wider trawl for a suitable model code be undertaken.

Selecting model codes

To decide if a model code is suitable for an application it is necessary to understand the code's characteristics and how they relate to the requirements of the application; the selection process then matches the model code to the application. While the modeller will bring knowledge about modelling techniques to this process, the water manager will bring knowledge of the application. Clearly both are required to make a decision about the model code to be used. Below, the characterisation of model codes is addressed, before considering how this helps the selection process.

As part of the Benchmark Models for the Water Framework Directive project (BMW, http://www.environment.fi/default.asp?contentid=116046&lan=EN) an inventory of model code characteristics was established to facilitate the model selection process. These stored characteristics are divided into three parts.

(1) *Generic characteristics*. These are basic facts including information about the domain (e.g. river, lake), computing platform (e.g. UNIX, Windows), cost basis on which the model code is made available, required user skills and documentation. Altogether there are seventeen items.
(2) *Model fact sheet.* These items are generally more descriptive than the generic characteristics, and include contact details for the model developer and references to publications describing the model code and model applications. Again, there are seventeen items.
(3) *Domain specific characteristics*. The domain-based nature of these characteristics ensures that they reflect the particular and different characteristics of the models. For the rivers domain, for example, there are 42 separate items including the modelled variables, mode (dynamic or probabilistic), number of modelled dimensions and the basis of process representation.

This information is generally factual and available in the public domain. A very few items require some interpretation; for example, modellers have different ideas about whether a model code that is based on widely accepted process understanding, but requires fitting to local conditions, should be described as 'process-based' or 'empirical'.

To decide if a model is suitable for a particular case study,

one could compare the requirements for the case study with the characteristics on the BMW model inventory (for river domain models a total of 76 characteristics). For each characteristic, an *a priori* statement would be needed to define what 'value' is acceptable for each characteristic. Of course, many characteristics would not be relevant and an assessment against them need not be made. If none of the model code characteristics is unacceptable, then the model code is suitable for the application. Such a comparison may indeed indicate whether or not a model is suitable but, when presented in this way, it does not sound like an attractive model code selection methodology. One reason for this is that the selection process appears to be driven by model characteristics. However, a more useful process, especially from the perspective of the water manager, would be to describe the requirement of the application and see how well a particular model code's characteristics match the requirement. In other words, the selection process should be driven by the requirements of the application. This is simply stated by Crabtree *et al*. (1986) as one of their two most important points in modelling criteria, i.e. "fitting the model to the problem, not the problem to the model". To do this demands that the requirements of the application are specified, and it could be argued that the characterisation of applications is more difficult than characterising model codes. To simplify this process requires a structure that enables the application specification to be no more detailed than necessary. There are some questions that will always need to be answered to define the application.

What is the management issue?
What are the key variables and processes?
What data are available?
What model outputs are required?
What solutions or understandings are expected?

While the first of these is a general question about the purpose of the study, the others all lead to a requirement for the model code. For example, if the key variables and processes have been identified, it is a requirement that they are included in the model code. The water manager would, therefore, answer the questions much as posed above, and the modeller would respond by stating whether a model code matches the requirement. While it is not explicit in the questions, by answering them the manager and modeller, based on the available information, have framed a conceptual understanding of the problem, the way it should be addressed and the outputs that are achievable.

It may well be that by answering these questions the water manager has specified the requirement completely but some additional requirements are likely. These might relate to

documentation,
sensitivity and uncertainty analysis,
potential for further development of the model application,
detail of process representation,
model parameterisation,
previous model usage, and
spatial and temporal scale and extent of the model application.

Where any of these is relevant, the water manager should specify the requirement and again, the modeller will respond by indicating if a proposed model code is suitable. However, even where the water manager indicates no particular requirement with regard to any of these considerations, the modeller would be well advised to make information about them available to the water manager. For example, where there is no specific requirement from the water manager regarding previous model usage, the modeller could indicate whether the model code has been previously used in other model applications, or is a new model code. Being open about issues of this type is likely to avoid any possible misunderstanding once the model application is under development.

A model selection protocol

A model selection protocol (MSP), strictly a *model code selection protocol*, has been developed based on the five questions given above, termed **mandatory factors**, and the eight further considerations, termed **optional factors**. The full text of the protocol is contained in Appendix 1, where the protocol is presented as a series of numbered questions referred to below as, for example, Q1.3. A key feature of the MSP is that it demands a dialogue between water manager and modeller. This is in contrast to the approach of Saloranta *et al*. (2003) which aims to "provide the growing number of (non-specialist) model users the means by which to judge from the model user's perspective … the appropriateness of a given model". Models should be applied by those with the necessary skills (i.e. a specialist), and the specialist must be involved in selecting the model code.

Mandatory questions relate to the most important aspects of the application and require the water manager to specify requirements. In evaluating a response to each question, the modeller selects the most appropriate single answer to indicate if the model code is suitable, unsuitable, or that further discussion is required. For example, in Q1.3, if all required data are available, the model code is suitable; if many essential data are missing the model is rejected, but between these extremes there may be a need to discuss further whether the model code should be used or not. A

model code marked as unsuitable can be immediately rejected from further consideration.

Optional factors allow the water manager to supply additional specifications where appropriate. For example, these can relate to spatial and temporal aspects of the modelling, parameterisation, and complexity. Where the water manager has specified a requirement, the evaluation is the same as for the mandatory factors, i.e. suitable, unsuitable, or further discussion required. Where this is not the case, the modeller checks a box describing the characteristics of the model code. For example, Q2.4 relates to model code parameterisation. If the water manager has a specific requirement, the modeller makes a direct evaluation against this on a yes/no basis. If the water manager has no specific requirement, the modeller indicates the nature of the parameterisation within the model code from a choice of four options, but there is no unsuitable answer. On completion of the evaluation, a model code may be considered suitable if satisfactory answers are given to all questions.

In all cases, the questions have been worded simply to avoid the confusion or ambiguity present in many of the criteria proposed by Saloranta *et al.* (2003). For example, they score a model as inadequate if its source code is unavailable and it is not being developed, yet these are probable characteristics of mature, reliable and stable model codes. Furthermore Saloranta *et al*. (2003) adopt a complex scoring and indicator system; for example the question concerning sensitivity and uncertainty requires three indicators from a list of five to be satisfied to achieve a good score, one from a list of three items to achieve an adequate score, or one from a list of two to be rated inadequate. One wonders from which end of the scale this assessment should start, and it would seem a distinct possibility that the score is indeterminate.

In most instances the modeller will have a particular expertise in a limited number of models. It is to be hoped that the modeller would quickly evaluate the information provided by the water manager and use this expertise to reject from this short-list model codes that are clearly unsuitable (i.e. the modeller undertakes a preliminary screening of model codes). The result of this is that only models that are likely to be satisfactory are submitted to an evaluation using the MSP. However, if the outcome of the evaluation is positive in the vast majority of cases, will the MSP be seen as an objective procedure for model code selection, or will it be taken as a means of "rubber-stamping" a decision that has already been taken?

To answer this question it is necessary to understand the reason for having confidence in the outcome of the selection. Perhaps the confidence comes if it is seen that some candidate model codes are rejected. For example if ten models were evaluated, and only one was judged suitable, the process would be seen to be discriminatory. But in fact the confidence in the validity of the discrimination must come from the knowledge that it is based on the right questions. If this is the case, and the right questions are being asked, then there is no need to identify unsuitable model codes to be confident in the positive assessment of those judged suitable. In other words, confidence comes from knowing that the right questions are being addressed. The MSP ensures this by highlighting the issues that should be considered by both water manager and modeller, inviting the former to specify the requirements and the latter to address them. On this basis the water manager should be pleased to be presented with a reliable assessment of why one model is suitable, rather than a longer report on why several others are not.

Of course there is nothing to say that the MSP should not to be applied to several model codes, but in doing so it should be understood that the outcome is unlikely to be that just one model emerges as suitable. It is probable that a number of model codes with different strengths and weaknesses are all found to be suitable, and there remains the problem of choosing between them.

One issue not addressed by the MSP relates to the resource implications associated with particular model codes. The authors agree with US EPA (1987) that "other general aspects of Model Selection Criteria which should never override objective, technical, or implementation criteria include budget, and staff and equipment resources". The MSP does, however, give some insight into what might be the non-standard elements within the modelling exercise. Mandatory questions 1.2–1.5, and optional questions 2.3–2.5, 2.7 and 2.8, all contain answers that indicate that the water manager's specification will require some enhancement of the model code. Clearly, this customisation of the basic model code to the particular requirements of the application is likely to have resource implications.

Testing the model selection protocol

A true test of the MSP might involve the following steps:

(1) A water manager and modeller use the protocol to evaluate the suitability of a number of model codes to achieve a specific management objective;
(2) Regardless of the outcome from 1, the model applications would be developed;
(3) The actual performance of the model could then be compared with the assessment from MSP.

In this way, how well the use of the protocol predicts the

actual performance of the model application might be judged by answering the following questions.

- Were the model codes deemed inadequate by the protocol actually unusable or unreliable in their application?
- Were those model codes assessed as suitable or all applicable in practice?

In some instances the answers will be straightforward, e.g. the absence of essential data would prevent a model code being applied. However, in other cases, answering these questions would be far from straightforward. For example, a model application may produce a result, but how is it decided that the result is reasonable? This would require some method of assessing model application performance against objective criteria and this leads to the second part of the model benchmarking process as envisaged by the BMW project, i.e. model assessment. For the present discussion, it is assumed that an assessment of performance can be made.

It is clear that such testing would require considerable effort to encompass a range of application types, and that a meaningful number of model codes would be used for each application. One way of reducing this effort would be to evaluate the criteria retrospectively against published model applications. However, since few, if any, publications describe failed model applications, this could not test completely the usefulness of the protocol to distinguish between adequate and inadequate models. Furthermore, many of the publications in this area are academic studies, rather than real-world applications; hence, the generic conclusion is that, in practical situations, model choice depends on the circumstances and objectives of the particular study.

While it appears impractical to evaluate fully the model assessment protocol, some tests can be made. For example, if used by several assessors, will the protocol yield consistent results, and would the resulting assessment be seen as a sound basis for the model application? These more modest objectives form the basis of the test in which the MSP and the benchmark criteria of Saloranta *et al.* (2003) were used for case studies from the UK and the Netherlands. In each case, the users of the model code selection tools were modellers who had themselves defined the objectives of the case studies. While this does not correspond with the ideal of the water manager and modeller being different individuals, the MSP separates these as functions and, therefore, this testing is worthwhile.

TEST 1: GENERAL WATER QUALITY OF TWO ENGLISH RIVERS

Three assessors have each used the selection tools to assess the suitability of the in-stream water quality model code QUESTOR (Eatherall, 1998, Naden *et al*., 2001, Boorman, 2003) to simulate the water quality class of two rivers in England. The classification to be used was developed by the England and Wales National Rivers Authority and is known as the General Quality Assessment (GQA). It uses percentiles of DO, expressed as a percentage of saturation, BOD and NH_4. The GQA was required on a reach basis, and will be checked against classification based on monitored data. If the model application represents GQA class, satisfactorily, it can then be used to assess management options to improve the quality of water within the two rivers.

Using the benchmark criteria, one assessor judged the model unsuitable on the basis of there being insufficient documentation, or tutorial, to enable a new user to apply the model. The requirement to express the result in terms of GQA class could also lead to the rejection of many model codes since few present the model output directly in this form. A specialist would realise what is involved in this step and deduce whether it could be accomplished in post-processing of model output, but it is not clear if a non-specialist would reach the same conclusion. Therefore, using the benchmark criteria, this could lead to the rejection of the model code (if the assessment that this process is not "via clear, well known and well-established links"), whereas in the MSP this would be noted as a required enhancement to the model code that could be achieved by the modeller, with the exact requirement agreed with the water manager. The QUESTOR model was used successfully for this model application.

TEST 2: POLLUTION LOADS TO THE RIVER AMSTEL

The objective was to predict the emissions from urban systems to the open surface waters in the Amstel basin using a model code that would give information on where and what pollution loads were entering the river system. Only a limited amount of data and time were available to perform this application and the model code needed to be chosen based on these requirements. The assessors were asked to review the selection of model code Sobek-EM (Emission Module). Sobek-EM is a simple book-keeping model code that enables the modeller to make a quick overview of the sources of emissions and the loads throughout the years. SOBEK has been widely used with many published applications, for example, Stelling *et al.* (1998), Verwey (2001), Dhondia and Stelling (2002) and Heeringen *et al.* (2002).

Using the MSP resulted in consistent responses from the users, whereas those from benchmark criteria were inconsistent. In fact, whereas all three assessors thought the model code suitable when using the MSP, only one reached the same conclusion using the benchmark criteria. Detailed knowledge of the model code and its development history, was one factor that made a difference in the assessments. The SOBEK-EM code was used successfully for this application.

When the assessors considered a different management objective within the same case study area, the SOBEK-EM model was rejected using the MSP; this application dependency, as expressed in the water manager's requirements, was seen as a useful feature of the MSP.

A wider context for the model code selection process

The MSP outlined in this paper is presented as a straightforward means of ensuring that an appropriate model code is used for a particular application. One reason for the need for such a protocol was presented at the outset, i.e. many model codes exist and it is necessary to ensure that a suitable one is chosen for a particular application. The MSP achieves this by recognising the expertise of the modeller to propose a model code that matches the requirements specified by the water manager.

But information describing model codes and applications is easily accessible to the water manager and other non-expert modellers. It appears in printed form as refereed publications, magazine articles, grey literature, advertisements and catalogues, and in a similar variety of forms on the internet. Rarely will any of this information make negative remarks about a model code or application; if problems have been encountered they will have been overcome. However, in recent years, model developers and model users have been encouraged to be mindful of the language they use in describing and to be more open about uncertainties associated with using models.

Model applications are often validated or verified but exactly what is meant by these terms is unclear. Oreskes *et al.* (1994) argue that numerical models of natural systems cannot be verified or validated, and suggest that the word confirmation is a better expression of the positive outcome of performance evaluation. Rykeil (1994), however, notes that these terms are synonymous in general usage, but have well understood technical usage; in other words they might be misunderstood by someone outside the modelling community but not by the experts within it. Also, for example, Lane and Richards (2001) noted that the language used to describe performance assessment is not neutral. While the term validation is used frequently, rarely is the process described as falsification; in other words by labelling the process validation assumes a positive outcome. Beven (1993) notes that while the validation of models is impossible, invalidation should be possible.

Carstensen *et al.* (1997) propose a terminology for modelling in support of water quality management, a discussion joined by Refsgaard and Henriksen (2004) who also extend the terminology to identifying roles of 'key players'. It remains to be seen if these suggestions receive widespread acceptance.

However, even with agreed terminology, there remains the problem of uncertainty in modelling. This uncertainty has many sources, notably, data (Høybye, 1996), parameter estimation (Freer and Beven, 1996), and model structure (Van Der Park, 1997; Thian Yew Gan *et al.* 1997). Beven (1993) introduces the concept of equifinality, i.e. that the same end might be achieved with not only different parameters within a given model structure but also different model structures. In other words, based on what is currently known and the data that are available, it is not possible to say which model structure is better. This acknowledgement of uncertainty and the ever increasing computing power available to modellers, encourages and enables the exploration of the uncertainty associated with modelling. Clearly, water managers need to be made aware of this uncertainty and to recognise that, while selecting a particular model code may help to reduce uncertainty, elimination of uncertainty is not possible. It is particularly important for a water manager to know if the uncertainties in the model estimates arising from the simulation of the current situation and the number of possible management scenarios are sufficiently small to allow the outcomes to be distinguished on a statistically significant basis.

It the context of this paper, models are being used to support management because there are few, or no, alternatives. As stated by Oreskes *et al.* (1994), it is the duty of the modeller to demonstrate the suitability, 'degree of correspondence', between the model and the real world. The MSP has been developed as a framework to enable such a demonstration of suitability. From the above discussion, it should be clear firstly, that many different models are suitable and, secondly, that despite any attempt to validate a model code, considerable uncertainty will be associated with the application.

Conclusions

The need for help in selecting models results from the large number of model codes that are available and the difficulty for the water manager in knowing that a particular code is

suitable for a given application. The problem is exacerbated by the use of potentially misleading technical terms in many model descriptions and the certainty, (meaning lack of uncertainty) with which they are presented. In practice, many models will be equally suitable for most applications but the water manager needs an auditable process to demonstrate that a suitable model code has been selected. Little advice on the whole of the model selection process is available in the open literature.

The model selection protocol enables a structured dialogue between a water manager and modeller to facilitate the selection of a model code for a particular application.

Key features of the MSP are that:

(1) it is application specific;
(2) it recognises the different roles in using models to support water management, and the expertise associated with those roles;
(3) it requires the water manager to specify the requirements of the application;
(4) it enables the modeller to state why a particular model code matches those requirements, and to present other characteristics of the model;
(5) it provides an audit trail for the decision to use a particular model code;
(6) it recognises that by correctly specifying model application requirements, the task is to identify one suitable model and not to screen a large number of possible models.

The MSP does not address issues relating to terminology and uncertainty specifically but does provide cues that can be used by both water manager and modeller. The proposed dialogue invites the modeller to provide this information even when no specific requirement is specified by the water manager. The MSP is not presented as a rigid protocol but as a protocol that can be adapted as required by water manager and modeller.

Acknowledgements

The authors gratefully acknowledge financial support for the project *Benchmark Models for the Water Framework Directive* from the European Commission (EK1-CT2001-00093), and the core science programmes of the Centre for Ecology and Hydrology and WL|Delft Hydraulics.

References

Bantz, I.B. and Cembrowicz, R.G., 1983. A combined dynamic flow and quality model for time variable effluents. *IAHS Publication no. 147,* 395–405.

Bart, J., 1995. Acceptance criteria for using individual-based models to make management decisions. *Ecol. Appl.,* **5**, 411–420.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. *Adv. Water Resour.*, **16**, 41–51.

Boorman, D.B., 2003. LOIS in-stream water quality modelling. Part 1: Catchments and methods. *Sci. Total Envir.*, **314/316**, 379–395.

Carstensen, J., Vanrolleghem, P., Rauch, W. and Reichert, P., 1997. Terminology and methodology in modelling for water quality management - a discussion starter. *Water Sci. Technol.*, **36**, 157–168.

Crabtree, R.W., Cluckie, I.D., Forster, C.F. and Crockett, C.P., 1986. A comparison of two river quality models. *Water Res.,* **20**, 53–61.

Dhondia, J. and Stelling, G.S, 2002. *Application of One Dimensional - Two Dimensional Integrated Hydraulic Model for Flood Simulation and Damage Assessment*. Fifth Int. Conf. on Hydroinformatics, 2002, Cardiff, UK.

Eatherall, A., Boorman, D.B., Williams, R.J. and Kowe, R., 1998. Modelling in-stream water quality in LOIS. *Sci. Total Envir.*, **210/211**, 499–518.

Emmerich, W.E., Woolhiser, D.A. and Shirley, E.D., 1989. Comparison of lumped and distributed models for chemical transport by surface runoff. *J. Environ. Qual.*, **18**, 120–126.

Freer, J. and Beven, K.J., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of dats: an application of the GLUE approach. *Water Resour. Res.*, **32**, 2161–2173.

Thian Yew Gan, Dlamini, E.M. and Biftu, G.F., 1997. Effects of model complexity and structure, data quality and objective functions on hydrologic modeling. *J. Hydrol.,* **192**, 81–103.

Heeringen, K.J. Van, Verwey, A. and Melger, E., 2002. Dutch Approach to high speed urban drainage modelling with SOBEK. *Proc. 9th Int. Conf. on Urban Storm Drainage*, 8-13 September 2002, Portland, USA.

Høybye, J.A., 1996. Uncertainty analysis in water quality modelling - Case study:Hjarbaek Fjord. *Nord Hydrol*., **27**, 203–214.

Houghton-Carr, H.A., 1999. Assessment criteria for simple conceptual daily rainfall-runoff models. *Hydrolog. Sci. J.,* **44**, 237–262.

Hughes, DA., 1994. Soil moisture and runoff simulations using four catchment rainfall-runoff models. *J. Hydrol.,* **158**, 381–404.

Lane, S.N. and Richards, K.S., 2001. The 'validation' of hydrodynamic models: some critical perspectives. In: *Model Validation: Perspectives in Hydrological Science,* M.G. Anderson and P.D. Bates (Eds.), Wiley, Chichester, UK. 413–439.

Moll, A. and G. Radach, G., 2003. Review of three-dimensional ecological modeling related to the North Sea shelf system. Part 1: models and their results. *Progr. Oceanography*, **57**, 175–217.

Naden, P.S., Cooper, D.M. and Boorman, D.B., 2001. Modelling large-scale river basins. In: *A Land Ocean Interaction Study: measuring and modelling fluxes from rivers to the coastal ocean*, D. Huntley, G.J.L. Leeks and D.E. Walling (Eds.), International Water Association, London, UK. 105–142.

Nix, J.S., 1990. Mathematical modelling of the combined sewer system. In: *Controls and treatment of combined sewer overflows,* P.E. Moffa (Ed.), Van Nostrad Reinhold, New York, USA. 23–78.

Oreskes, N., Shrader-Frechette, K. and Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263**, 641–646.

Refsgaard, J.C. and Henriksen, H.J., 2004. Modelling guidelines – terminology and guiding principles. *Adv. Water Resour.*, **27**, 71–82.

Refsgaard, J.C., Henriksen, H. J., Harrar, B., Scholten, H. and Kassahun, A., 2005. Quality assurance in model based water management - review of existing practice and outline of new approaches. *Environ. Model. Software*, **20**, 1201–1215.

Rykiel, E.J., 1994. Comment on Oreskes *et al*. *Science*, **264**, 330–331.

Saloranta, T.M., Kamari, J., Rekolainen, S. and Malve, O., 2003. Benchmark criteria: a tool for selecting appropriate models in the field of water management. *Environ. Manage.*, **32**, 322–333.

Series on Model Choice, Cooperative Research Centre for Catchment Hydrology, www.toolkit.net.au/modelchoice.

Shoemaker, L., Lahlou, M., Bryer, M., Kumar, D. and Kratt, K., 1997. *Compendium of tools for watershed assessment and TMDL development.* US EPA Washington DC, EPA841-B-97-006.

Stelling, G.S., Kernkamp, H.W.J. and Laguzzi, M.M., 1998. Delft Flooding System: a powerful tool for inundation assessment based upon a positive flow simulation. In: *Hydroinformatics*, Babovic and Larsen, (Eds.), Balkema, Rotterdam, The Netherlands. 449–456.

US Environmental Protection Agency, 1987. *Selection criteria for mathematical models used in exposure assessments: Surface Water Models.* US EPA, Washington DC, EPA/600/8-87/042.

Van Der Molen, D., 1999. *The role of eutrophication models in water management*. PhD Thesis, Wageningen University. RIZA report 99.020, Lelystad, The Netherlands.

Van Der Perk, M., 1997. Effect of model structure on the accuracy and uncertainty of results from water quality models. *Hydrol. Process.*, **11**, 227–239.

Vandewiele G.L. and Ni-Lar-Win, 1998. Monthly water balance models for 55 basins in 10 countries. *Hydrolog. Sci. J.,* **43,** 687–699.

Van Oijen, M., Rougier, J. and Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol.,* **25**, 915–927.

Van Waveren, R.H., Groot, S., Scholten, H., Van Geer, F.C., Wösten, J.H.M. and Koeze, R.D., 1999. *Good modelling practice handbook.* STOWA Report 99-05. Utrecht, RWS-RIZA, Lelystad, The Netherlands.

Verwey, A., 2001. Latest developments in floodplain modelling - 1D/2D integration. *Proc. 6th Conf. on Hydraulics in Civil Engineering*, The Institution of Engineers, Hobart, Australia.

Warwick, J.J. and Roberts, L.A., 1992. Computing the risks associated with wasteload allocation modelling. *Water Res. Bull.,* **28**, 903–915.

Wasson, J.G., Tusseau-Vuillemin, M.H., Andréassian, V., Perrin, C., Faure, J.B. and Barreteau, O., 2003. What kind of water models are needed for the implementation of the European Water Framework Directive? Examples from France. *Int. J. River Basin Manage.*, **1,** 125–135.

Appendix 1 Model selection protocol

MANDATORY FACTORS

The modeller's evaluation informs the end user of the suitability of the selected model to undertake the management task. Where the result is a tick ✓ the model is suitable, a cross **x** indicates it is not-suitable and the model under consideration can be rejected immediately; a **?** implies further discussion is required between the water manager and modeller.

	Water manager's specification	*Modeller's evaluation*
MANAGEMENT OBJECTIVES		
1.1 What are the management objectives?	Management Objectives (in brief)	Is the model code relevant to these management objectives? ● Yes ✓ ● No **x**
REQUIREMENTS FOR MODEL FUNCTIONALITY		
1.2 What, in general terms, is the model functionality (e.g. key variables and processes) that is required to support the management objective?	Model functionality relevant to management objectives	● The model code directly includes the required functionality. ✓ ● It is a straightforward task to include the required functionality in the model code and this can be undertaken by the modeller. ✓ ● To include the required model code functionality appears straightforward but requires the involvement of a third-party model developer. **?**

	Water manager's specification	*Modeller's evaluation*
		● It is not possible to use this model code to generate the required outputs. **x**
DATA AVAILABILITY **1.3** Describe the data available for the model application.	Data availability	Based on the described data availability, can the model application be expected to produce the required model application outputs in support of the management objectives? ● Yes, all required data are available. ✓ ● Yes, but some surrogate data will have to be used where essential data are missing. ✓ ● Yes, but this will rely on many assumptions regarding missing data **?** ● No, too many key data are missing to make the model application worthwhile. **x**
REQUIREMENTS FOR MODEL OUTPUTS **1.4** What, in general terms, are the model-derived outputs that are required to support the management objective?	Model outputs relevant to management objectives	● The model code directly produces the required outputs. ✓ ● It a straightforward task to produce the required outputs from the model output and this can be undertaken by the modeller. ✓ ● It is a straightforward task to interpret model application outputs according to the management objective and this can be presented by the modeller. ✓ ● It is not possible to use this model code to generate the required outputs. **x**
INTERACTION WITH THE MODEL **1.5** What further interaction with the model is required e.g. for further scenario analysis, presentation to stakeholders?	Requirements for model interaction	● The model code allows the required interaction but interaction, and this can be done by the end-user. ✓ ● The model code allows the required interaction but would need the involvement of the modeller. ✓ ● It is not possible to interact with the model code in the way required. **x**

OPTIONAL FACTORS

The meaning of the evaluation is as for the mandatory factors, with the exception that empty boxes ¨ are for the modeller to check as a means of providing feedback to the water manager.

	Water manager's specification	*Modeller's evaluation*
MODEL SPATIAL RESOLUTION AND SCALE **2.1a** Do you have particular requirements relating to the spatial resolution of the adopted modelling approach?	● Yes What is the envisaged spatial resolution for the model application? (use a range of values if appropriate) ● No	Can the model code operate at the desired spatial resolution? ● Yes ✓ ● No **x** Can the model operate at spatial resolution that is appropriate for the management objective and to produce the required outputs? ● Yes ✓ ● No **x**

	Water manager's specification	*Modeller's evaluation*
2.1b Do you have particular requirements relating to the spatial extent scale (extent) of the adopted modelling approach?	● Yes What is the envisaged spatial scale for the model application? (use a range of values if appropriate) ● No	Can the model code operate at the desired spatial scale? ● Yes ✓ ● No **x** Can the model operate at a spatial scale that is appropriate for the management objective and to produce the required outputs? ● Yes ✓ ● No **x**
MODEL TEMPORAL RESOLUTION AND SCALE		
2.2a Does the model operate in a dynamic way that makes it appropriate to consider temporal resolution and scale of the inputs and/or outputs?	● Yes Answer questions 2.2b and 2.2c. ● No Answer question 2.2d.	
2.2b Do you have particular requirements relating to the temporal resolution of the adopted modelling approach?	● Yes What is the envisaged temporal scales for the model application? (use a range of values if appropriate) ● No	Can the model code operate at the desired temporal resolution? ● Yes ✓ ● No **x** Can the model code operate at a temporal resolution that is appropriate for the management objective and to produce the required outputs? ● Yes ✓ ● No **x**
2.2c Do you have particular requirements relating to the temporal scale of the adopted modelling approach?	● Yes What is the envisaged temporal scale for the model application? (use a range of values if appropriate) ● No	Can the model code operate at the desired temporal scale? ● Yes ✓ ● No **x** Can the model code operate at a temporal scale that is appropriate for the management objective and to produce the required outputs? ● Yes ✓ ● No **x**
2.2d Specify the mode in which the model code will operate (e.g. steady state, stochastic)?	Mode of operation	Can the model code operate at the desired mode? ● Yes ✓ ● No **x**
MODEL COMPLEXITY (PROCESS REPRESENTATION)		
2.3 Do you have particular requirements relating to the complexity of process representation within the model for use in the application?	● Yes Complexity requirements	● The model code operates at a r range of complexities that encompass the required level of complexity. ✓ ● It is a straightforward task for the modeller to make the model code operate with the desired level of complexity. ✓ ● The model can only operate at a level of complexity that exceeds the specified requirements. **?** ● It is not possible to operate at the specified level of complexity. **x**

	Water manager's specification	*Modeller's evaluation*
	● No	Does the model code encompass the level of complexity appropriate to the management objective (i.e. is it neither too simple not too complex)? ● Yes ✓ ● No **x**
MODEL PARAMETERIZATION		
2.4 Do you have specific requirements regarding model parameterisation (e.g. identifiability)	● Yes Specific requirements ● No	Does the model code achieve the specific requirements? ● Yes ✓ ● No **x** ● The model code has no parameters that require application specific values. ☐ ● The model code requires application specific parameters but these can be derived from published values. ☐ ● The model code requires application specific parameters and these are set by expert knowledge of the modeller. ☐ ● The model code requires application specific parameters and these are set by calibration to be undertaken by the modeller. ☐
MODEL SENSITIVITY		
2.5 Do you have specific requirements regarding model sensitivity?	● Yes Specific requirements ● No	Does the model code achieve the specific requirements? ● Yes ✓ ● No **x** Tick as appropriate. ● Sensitivity analysis is built into the model code and will be undertaken by the modeller. ☐ ● Sensitivity analysis can be undertaken externally to the model code by the modeller. ☐ ● Sensitivity analysis has been published for this model code. ☐ ● No sensitivity is to be undertaken as part of the study. ☐
MODEL VALIDATION AND USAGE		
2.6 Do you have specific requirements regarding model formulation, validation and usage?	● Yes Specific requirements ● No	Does the model code achieve the specific requirements? ● Yes ✓ ● No **x** Tick as appropriate. ● The published information is sufficient to have confidence in the validity of the model code. ☐ ● There have been previous applications of the model code and sufficient to have confidence in applying the model, but they are not widely published. ☐

	Water manager's specification	*Modeller's evaluation*
		● This is a new model code or model code adaptation without previous application in this form. ☐
MODEL DEVELOPMENT		
2.7 Do you have particular requirements relating to the further development of the model application?	● Yes Development requirements	● The model code already contains the required functionality to extent the proposed application. ✓ ● It is a straightforward task for the modeller to enhance the model code to enable the outlined developments. ✓ ● To enhance the model code to enable the outlined developments will require the model developer. ✓ ● It is not possible to further develop the application according to these requirements. **x**
	● No	Is the model code already at the limit of its capabilities in meeting the existing requirements application? ● Yes ☐ ● No ☐
MODEL DOCUMENTATION		
2.8 Do you have particular requirements relating to model documentation?	● Yes Documentation requirements	Does the model documentation achieve the specific requirements?· Yes ✓ ● Additional documentation will be prepared by the modeller to meet the specific requirements. ✓ ● No **x**
	● No	Is the model documentation as necessary to undertake the modelling, and adequate to describe model function? ● Yes ✓ ● Additional documentation will be prepared by the modeller. ✓ ● No **x**