Decision Support Systems in Agriculture, Food and the Environment:

Trends, Applications and Advances

Basil Manos
Aristotle University of Thessaloniki, Greece

Konstantinos Paparrizos
University of Macedeonia, Greece

Nikolaos Matsatsinis
Technical University of Crete, Greece

Jason Papathanasiou
University of Macedeonia, Greece

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey · New York

Director of Editorial Content: Kristin Klinger
Director of Book Publications: Julia Mosemann
Acquisitions Editor: Lindsay Johnston
Development Editor: Christine Bufton
Publishing Assistant: Travis Gundrum; Natalie Pronio
Typesetter: Keith Glazewski; Travis Gundrum; Casey Conapitski
Production Editor: Jamie Snavely
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Decision support systems in agriculture, food and the environment : trends, applications and advances / Basil Manos ... [et al.], editors.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-61520-881-4 (hbk.) -- ISBN 978-1-61520-882-1 (ebook) 1. Agriculture--Decision making. 2. Agriculture--Data processing. 3. Agriculture--Management--Information technology. 4. Agricultural informatics. 5. Decision support systems. I. Manos, Basil, 1950-
S494.5.D3D39 2010
630.685--dc22
2009052438

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 10
Validation of a Model Appropriateness Framework Using the Elbe Decision Support System

Yue-Ping Xu
Zhejiang University, China

Martijn J. Booij
University of Twente, The Netherlands

ABSTRACT

This paper describes validation of an appropriateness framework, which has been developed in a former study, to determine appropriate models under uncertainty in a decision support system for river basin management. Models are regarded as 'appropriate' if they produce final outputs within adequate uncertainty bands that enable decision-makers to distinguish or rank different river engineering measures. The appropriateness framework has been designed as a tool to stimulate the use of models in decision-making under uncertainty and to strengthen the communication between modelers and decision-makers. Through the application to a different river with different objectives in this validation study from the river used in the development stage, this paper investigates whether the appropriateness framework works in a different situation than it was designed for. Recommendations from the development stage are taken into account in this validation case study as well. The final results from the study showed a successful validation of the appropriateness framework and suggested further possibilities for the application in decision support systems for river basin management.

INTRODUCTION

In river basin management, often a range of models exist which can be used to describe underlying physical, socio-economic and ecological processes of interest. The availability of complex models as management tools is growing as the power of computers increases. Models are normally selected based on the conditions that they 'best' fit a particular set of data (Forster, 2000; Wasserman, 2000; Boorman, 2007). However, this condition

DOI: 10.4018/978-1-61520-881-4.ch010

may not be the only concern when people choose a model especially for management purpose.

Although, until now, there has been no single standard definition of model complexity, Brooks and Tobias (1996) defined model complexity as a measure of the number of constituent parts and relationship in the model. Busemeyer and Wang (2000) argued that complex models often have an excessively large number of parameters. Recently, there are many arguments about the use of complex and simple models (Jakeman & Hornberger, 1993; Nihoul, 1994). In general, modelers often have the intention to develop more complex models than models actually used in river basin management. These complex models help getting a better understanding of the system to be modeled or answering more complex problems, and can provide a good reference background to simpler models used in management. Decision-makers, however, focus more on practical applications of models to solve their management problem and prefer simpler models (Vreugdenhil, 2006). Parker et al. (1995) argued that complex and sophisticated models can be easily misused. The more variables in the model, the more difficult it becomes to use as a practical management tool. Moreover, for complex models, their outputs often have no measure of confidence associated with them. Principles for using a model in a planning study or strategic management are somewhat different from those for model development. According to Vreugdenhil (2002), a model for planning purposes often needs to provide only integrated, not very detailed information.

Several studies of reservoir, hydrologic, flood routing and water quality models have demonstrated that simpler model formulations are often more accurate than more complex formulations (Loague & Freeze, 1985; Palmer & Cohan, 1986; Jakeman & Hornberger, 1993). Robinson and Freebairn (2001) also made an interesting observation: very common among the conclusions of papers at MODSIM (International Congress on Modelling and Simulation) are expressions on the need for future improvements of the models in order to make them realistic. It is much less common to find conclusions suggesting that a problem has been solved or that models can be simpler for management purposes. In the case of decision support systems (DSS), many data, knowledge and models are put together whereas some developers of DSS prefer to use models as complex as possible, which often makes the system hard to understand, use, maintain and, moreover, may cause considerable uncertainty. For decision support systems, it is argued that often the complexity of models largely exceeds the actual requirements.

To stimulate better use of models in decision-making and increase the reliability of decision-making under uncertainty, the use of appropriate models in a DSS for river basin management has been proposed by Xu et al. (2007). They argued that, for a model-based DSS, models are supposed to fit decision-makers' use but without leaving out essential mechanisms associated with management problems. Models are regarded as appropriate if they produce final outputs within adequate uncertainty bands that enable decision-makers to distinguish or rank different river engineering measures. The ranking of measures is the management problem decision-makers aim to solve and uncertainty analysis plays an important role on determining appropriate models in a DSS. Based on these concepts, a systematic approach – *appropriateness framework* – to determine appropriate models under uncertainty in decision support systems for river basin management has been developed (Xu et al., 2007). They have also used the Dutch Meuse DSS as a development case study to demonstrate the appropriateness framework and have succeeded in deriving appropriate models. At the end, they made a few recommendations based on the development study for further applications of the framework in other DSSs.

This paper hence focuses on the validation of the appropriateness framework developed by Xu et al. (2007). The goal of validation is to investigate whether or not the appropriateness framework works in a different situation than it was designed for. To validate the framework while taking into account the recommendations from the development case study, a second DSS is thus needed. The requirements for the second case study are that: (i) it should be a different river with different objectives; (ii) adequate data should be available to apply the framework; and (iii) it should be possible to establish good contacts with decision-makers. After taking above requirements into account, the Elbe Decision Support System is chosen because this case study is the best option available (Matthies et al., 2003). Besides, to ensure the whole appropriateness framework is properly validated, two validation criteria are proposed: (i) it should be able to derive appropriate models at the end; and (ii) the framework should be efficient to derive appropriate models.

The contents of this paper are as follows. First of all, the appropriateness framework and its main steps are described. Second, a brief description of the validation case study is made. Then, the results from applications of each step of the framework are given. Finally, relevant conclusions based on the Elbe DSS case study are drawn and discussion is made.

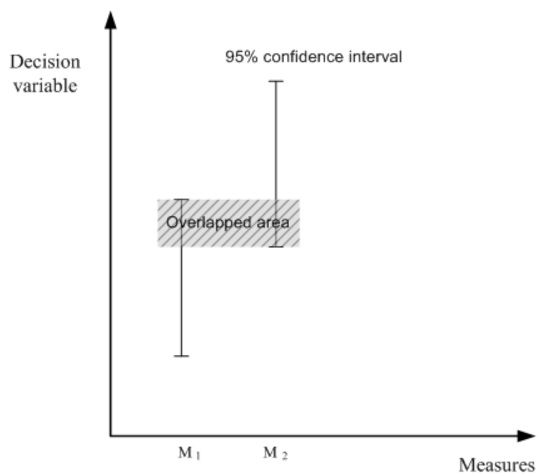
APPROACH: APPROPRIATENESS FRAMEWORK

General Introduction

The determination of appropriate models in a DSS largely depends on the definition of what is appropriate. As mentioned earlier, models are regarded as appropriate if they produce final outputs within adequate uncertainty bands that enable decision-makers to rank different measures. However, it is already noted that, the determination of appropriate models generally not only depends on the practical ranking problem under uncertainty, but also on other aspects such as user-friendliness, flexibility of models and computation time. Basically, good practice is the platform for pursuit of model quality or appropriateness (Jakeman et al., 2006).

In the paper of Xu et al. (2007), the ranking problem and the uncertainty are regarded as the most critical aspects in determining the appropriateness of models. The ranking problem is briefly introduced in Figure 1, where two measures M_1 and M_2 are used for illustration. The error bars are 95% confidence intervals of model outputs from the two measures. In most cases, M_2 generates higher outputs than M_1 (the probability is more than 50%). If a higher value indicates a better measure, M_2 is better than M_1 in most cases. However, M_1 could be better because of the uncertainty involved. The overlapped area in this figure indicates where the ranking of M_1 and M_2 can be different from the ranking in most cases. Therefore, the key is how to obtain an acceptable ranking of measures under uncertainty, viz. when can the ranking of M_2 being better than M_1 be accepted? The proposed criterion is designed as the risk of obtaining an unacceptable ranking. This criterion is defined in such a way to help rank measures under uncertainty and is defined with reference to the classical concept of risk, which usually considers the probability of a hazard and the consequences of that hazard (Kaplan & Garrick, 1981). It takes into account two important aspects in which decision-makers are interested. The first aspect of this risk is the mean difference of model outputs among different measures. The second aspect is the probability of obtaining an unacceptable ranking which is the probability for decision-makers to obtain a different ranking from the ranking in most cases. By combining these two aspects, the risk of obtaining an unacceptable ranking is defined as the product of the mean difference and the corresponding probability: $R = \overline{Y} * P$, where $\overline{Y}$ is the mean difference and P is the probability of obtaining an unacceptable ranking. The value of the mean

Figure 1. Brief description of the ranking problem



difference is important to decision-makers indicating which measure is better on average and regarded as the loss when an unacceptable ranking is adopted. Moreover, the mean difference is used as a scaling factor. The risk utilizes the mean difference to scale the effect of probability and is independent of measure with different levels of effects on the system. The method to calculate the risk of obtaining an unacceptable ranking can be found in Xu et al. (2007).

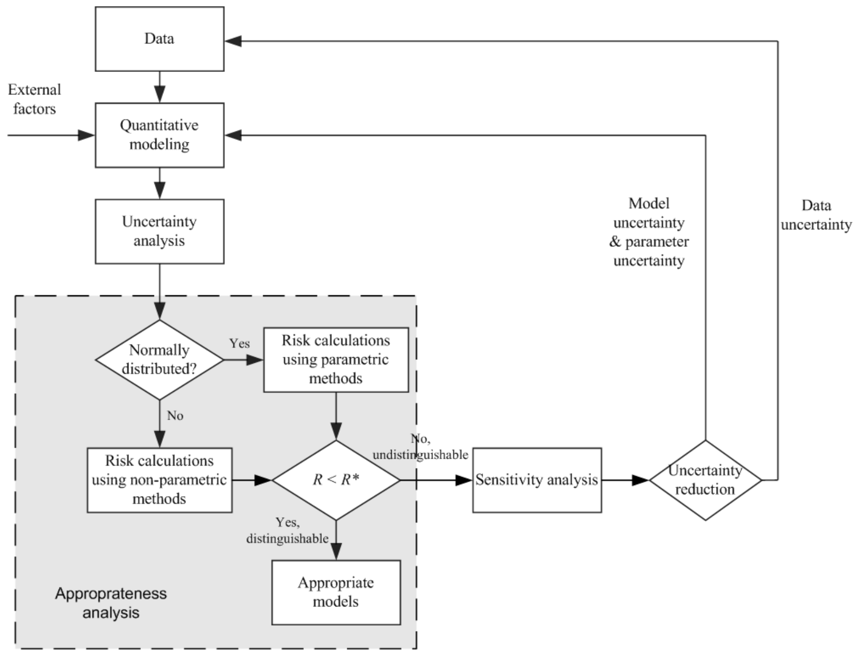
If the risk calculated can be accepted by decision-makers, the ranking of measures can be accepted and models are regarded as appropriate. Therefore, to determine appropriate models in a DSS, an acceptable level of this risk needs to be determined. This value is often obtained based on the requirements or experiences of decision-makers. It is an indication of an acceptable balance of costs and benefits and the uncertainty in decision variables. Different from the objective risk, the acceptable risk is a subjective one. Different decision-makers may choose different acceptable risks. Questionnaires or interviews with relevant decision-makers can be used to determine this value. The models are considered to be appropriate if the risk calculated is smaller than this acceptable risk.

Main Steps of the Framework

Figure 2 displays the general appropriateness framework, which is adapted from the framework introduced in Xu et al. (2007). This framework aims to achieve appropriate models under uncertainty in a systematic way and is mainly designed for planning and strategic management purposes. Models are first chosen or built by modelers on the basis of decision-makers' problems and objectives (so called 'quantitative modeling'). The appropriateness framework generally starts from simple but reasonable models which include the most important and relevant processes of the system to be modeled. In this figure, R is the risk of obtaining an unacceptable ranking when considering a pair of river engineering measures and R^* is the acceptable risk defined by decision-makers through questionnaires or interviews. The main steps in this appropriateness framework include: (i) uncertainty analysis; (ii) appropriateness analysis; and if necessary (iii) sensitivity analysis; (iv) model improvement (uncertainty reduction). For the development, choice, or use of such appropriate models, expertise will play an important role.

According to Figure 2, after quantitative modeling, uncertainty analysis will be employed. The aim of uncertainty analysis in this framework is to identify the sources of uncertainty in the models and data and propagate them into the final model outputs. Various uncertainty analysis methods are available, including Monte Carlo simulation and Bayesian approach. Description of them can be found in Satelli et al. (2000). Appropriateness analysis aims to analyze whether or not the models used in the DSS are appropriate by calculating the risk of obtaining an unacceptable ranking. If the risk is smaller than the acceptable risk, the ranking can be accepted and models are considered to be appropriate. When multiple measures are considered, the models are regarded as appropriate

Figure 2. General appropriateness framework (R represents the risk of obtaining an unacceptable ranking and R^* represents the acceptable risk)



if the risks for all pairs of measures are acceptable. If models are judged to be inappropriate, sensitivity analysis is conducted to identify the major sources of uncertainty in the models and data, whose contributions to the total uncertainty in model outputs are large. If uncertainty reduction in these major sources is needed, various possible means to reduce the uncertainty can be implemented which include collecting more and/or high quality measurement data, increasing model complexity, increasing spatial resolution or time step of data, using better data processing methods and collecting expert opinions etc. However, when the cost of model improvement is too high or when improvement can no longer reduce the risk calculated, it may not be worthwhile to put more efforts on reducing the uncertainty. More detailed information about the techniques used in the framework can be found in Xu et al. (2007).

Recommendations from Development Case Study

Xu et al. (2007) made several recommendations for further applications of the appropriateness framework. These recommendations from the development case study will be investigated through the validation study. They include: (i) consideration of model uncertainty; (ii) realistic uncertainty reduction; (iii) realistic determination of acceptable risk; and (iv) consideration of interdependency among model outputs from different measures. Investigation of the recommendations in the validation case study can generalize the application of the proposed appropriateness framework in river basin management by emphasizing the differences between two case studies. In the following, adequate treatments of these recommendations are described.

Model Uncertainty

When applying the appropriateness framework to the Dutch Meuse DSS, model uncertainty was ignored, which may have an important contribution to the uncertainty in model outputs. Model uncertainty stems from assumptions in a mathematical model of a physical system, including governing equations, spatial resolution and time step used in solving the model. The usual way to incorporate model uncertainty in uncertainty analysis is Bayesian model choice or Bayesian model averaging (e.g. Draper, 1995; Gilks et al., 1996; Devooght, 1998; Wasserman, 2000; Neuman, 2004). Another often-used method to deal with model uncertainty is by comparing the model outputs from models with different model complexity (e.g. Loague & Freeze, 1985; Turner et al., 1996; Vreugdenhil, 2002; Booij, 2005) and then investigating how they affect the appropriateness analysis. In the validation case study of Elbe DSS, the latter method will be used because of its ease of use and consistency with the concept of appropriateness. In this paper, the effect of model uncertainty on the appropriateness analysis will mainly be investigated by improving mathematical equations of the most important components in the system identified by sensitivity analysis. If the risks calculated from models of different complexity are all acceptable to decision-makers, simpler models will be considered to be more appropriate.

Realistic Uncertainty Reduction

In the Dutch Meuse DSS case study, the reduction of uncertainty was implemented by two hypothetical cases, assuming a certain percentage for the reduction of uncertainty. In the Elbe DSS case study, more realistic reduction of uncertainty will be considered. The techniques mentioned before will be used to reduce the uncertainty, e.g., by collecting more or higher quality data or by increasing model complexity. It is expected that these uncertainty reduction techniques can reduce the uncertainty in model outputs and thus reduce the risks to a level acceptable to decision-makers.

Realistic Determination of Acceptable Risk

One important aspect in the appropriateness framework is to determine the acceptable risk used to analyze whether or not the models used in the Elbe DSS are appropriate. This value, representing the attitudes of decision-makers to uncertainty, plays a significant role in determining the efforts needed to achieve appropriate models. A questionnaire is hence designed to determine the mean difference of model outputs and the probability of obtaining an unacceptable ranking. Two questions respectively for the shipping and vegetation models are contained in the questionnaire. For the shipping model, the questions are: 1) what is the acceptable mean difference of the number of shipping days per year to get an acceptable ranking? 2) what probability is allowable to obtain an unacceptable ranking of measures as far as the numbers of shipping days per year are concerned? Thirty copies of these questionnaires were sent by email to two groups of participants. One group includes 10 decision-makers. The other group includes 20 expert colleagues experienced in modeling and uncertainty analysis. The choice for the second group is because of the difficulty to get enough contacts with decision-makers involved in the Elbe DSS. Another reason for this choice is that the acceptable risk itself is subjective and may change among decision-makers. It is expected that expert colleagues with good experience on modeling and uncertainty analysis may act as decision-makers.

Interdependency among Model Outputs from Different Measures

In the Dutch Meuse DSS case study, it was assumed that model outputs from different river engineering measures were independent of each other. This assumption is a simplification of reality made to keep the case study simple for design purposes, which mainly indicates that the uncertainties originating from the models and data are independent of each other (Reichert & Borsuk, 2005). Another extreme assumption could be that all uncertainties are fully dependent and then model outputs from different measures will be highly dependent. In the Elbe DSS case study, a more reasonable assumption is to assume some degree of interdependency, which means the correlation between model outputs should be taken into account based on the characteristics of different measures.

ELBE DSS VALIDATION CASE STUDY

Introduction

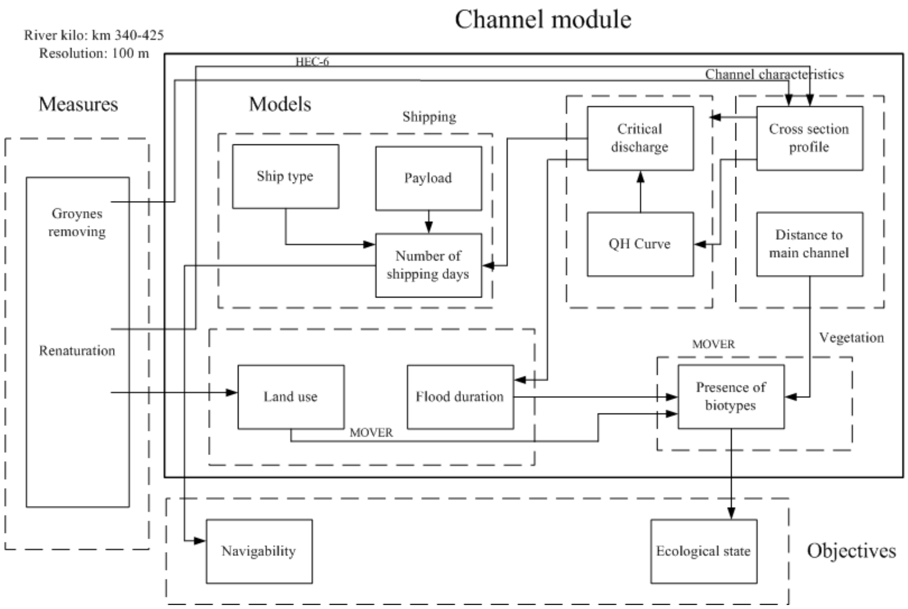
The Elbe DSS is used to investigate the effects of river engineering measures on the Elbe River which is one of the largest rivers in Central Europe, flowing from its source in the Czech Republic into the mouth at the North Sea in Germany (Matthies et al., 2003). The Elbe River has a length of about 1,100 km and its basin area is approximately 140,000 km^2. The German part of the river basin forms two third of the total river basin area and is inhabited by 20 million people (De Kok et al., 2000). The average discharge at the river mouth is about 700 m^3/s (Krysanova et al., 2004).

A full inventory of problems, problems owners, and objectives for the Elbe River can be found in De Kok et al. (2000). The main problems consist of lack of navigability along the Elbe due to low flow conditions, flood risk and lack of biodiversity in the floodplains of the Elbe River. The problem owners include German Institute of Hydrology, Water and Shipping Administration, Public Institution for Hydraulics, Elbe Nature Reservation Centre etc. The main objectives are therefore to maintain a minimum state of navigability along the Elbe River, to reduce flood risk along the Elbe, and to improve the biodiversity of the floodplains. Measures intended to solve these problems include dike heightening, renaturation, retention basin, dike shifting, groyne modification and channel dredging.

For the validation purpose of the appropriateness framework, the channel module contained in the Elbe DSS is chosen (De Kok et al., 2000). This module considers three objectives: ecological state of the floodplains, flood risk/safety and navigability along the Elbe main channel. The channel module focuses on the river section between the Czech Border (km 0) up to Weir Geesthacht (km 568), which is the outlet to the North Sea. The ecological state of the floodplains is described in terms of the dominant vegetation groups (biotypes), and depends primarily on the flood duration in the floodplains, land use, and distance to the main channel. Flood risk/safety is expressed in terms of relative damage (as compared to the maximum damage) and dependent on inundation depths in the floodplains for selected discharges. Shipping intensity along the Elbe is low due to unfavourable hydraulic conditions and numerous points where bed level changes make passages of ships difficult. In the Elbe DSS, navigability is expressed as the number of days per year a standard vessel can pass nine navigation sections distinguished in the German Elbe River (De Kok et al., 2000). Due to the availability of data and models, in this case study, only the ecological state and navigability functions are taken into account within the river section 340-425 km, with a total length of 86 km. Figure 3 shows a simplified system diagram for the Elbe DSS used for the validation of the appropriateness framework.

The decision variables in this case study are the number of shipping days per year for the function navigability and the frequency of dominant biotypes in the floodplains for the function ecological state. As shown in Figure 3, two measures are implemented to investigate the effects on decision

Figure 3. Simplified system diagram for the Elbe DSS



variables, namely, groyne removing in the main channel and renaturation in the floodplains. Only two measures are used herein because of the availability of data, models and the DSS development situation at the stage of this research.

Models

The models used to investigate the effects of two measures on different decision variables in the Elbe DSS include hydrological models, hydraulic models, the shipping model, and the vegetation model. They are described in the following.

Hydrological Models

In the Elbe DSS, the lognormal distribution is employed to model daily average discharge statistics and calculate the flood duration (the number of days per year a discharge $Q*$ is exceeded at location x). The cumulative probability of exceedance of a discharge $Q*$ for a lognormal distribution is given by:

$$P\left(Q > Q^{*}\right) = F(Q^{*}) = 0.5\left[1 - erf\left[\frac{\left(\log\left[Q^{*}\right] - \mu(x)\right)}{\sigma(x)\sqrt{2}}\right]\right] \quad (1)$$

Where $Q*$ is discharge; ¼$\left(x\right)$ and $\sigma(x)$ are mean value and standard deviation respectively. The error function is given as: $erf(z) = \frac{2}{\sqrt{\pi}}\int_0^Z dx\ e^{-x^2}$

Hydraulic Models

Two 1D hydraulic models are used to provide inputs to the shipping and vegetation models. One is based on the fitted rating curves and the other one is the HEC-6 model (US Army Corps of Engineers, 1993). The rating curves describe the relationship between the discharge and river stage (water level) along the river mainly based on measurements. The available rating curves are based on existing cross profiles of the river:

$$H = aQ^b \tag{2}$$

Where H is water level; Q is discharge; a and b are coefficients which can be estimated by regression analysis.

The main function of the HEC-6 model is to deal with the computation of water surface profiles for steady gradually varied flow in river channels. It is a one-dimensional open channel flow numerical model. The model calculates the water levels for 10 different discharges for every 100 m along the river section. These 10 different discharges correspond to discharges of different return periods, which represent a whole range of the river flow in the Elbe. The advantage of HEC-6 is its capability to take into account the effects of river engineering measures. One disadvantage is that this model cannot be used directly to calculate the flood duration in the Elbe DSS. A solution is therefore to combine both the HEC-6 model and the rating curves. First, the HEC-6 model is used to produce the water level and discharge data for different measures. Second, the rating curves are used to model the relationship between these discharges and water levels calculated from the HEC-6 model. By doing this, new rating curves can be obtained with the HEC-6 model to account for measures that affect the geometry of channel and floodplains, such as renaturation by changing the roughness of the floodplains. The water level and discharge data obtained from HEC-6 are used to derive the parameters a and b in Eq. (2) for each location along the concerned river section 340-425 km and each measure. The rating curves obtained will then be used as inputs to the shipping and vegetation models in the DSS.

Shipping Model

The purpose of the shipping model is to estimate the navigability of the concerned river section 340-425 km. The navigability is calculated for every location x along the river section where locations x are within a distance of 100 m of each other (sub-sections). The shipping model first calculates the critical discharge for the navigability for a given minimal water depth. With this critical discharge and the discharge statistics, the number of shipping days per year can be calculated for each river sub-section (every 100 m). The navigability is then represented by the 10 percentile of the number of shipping days per year along the concerned river section. The 10 percentile value indicates the value that is greater than 10 percent of all the data (the number of shipping days) along the concerned river section (340-425).

The number of shipping days per year at channel location x follows from the discharge statistics and the critical discharge value:

$$N_{nav}(x) = \frac{365}{2}\left(1 - erf\left[\frac{\log\left(Q_{nav}(x)\right) - \mu\left(x\right)}{\sigma\left(x\right)\sqrt{2}}\right]\right) \tag{3}$$

Where Q_{nav} is the critical discharge value and ¼$\left(x\right)$ and $\sigma(x)$ are mean value and standard deviation respectively. The number of shipping days per year in the concerned river sections is calculated as the 10 percentile of all the numbers of shipping days per year (every 100 m).

Vegetation Model (MOVER)

The vegetation model is used to produce maps for the dominant groups of vegetation (biotypes) in the floodplain area. The water levels in the floodplain are assumed to be the same perpendicular to the river. The large-scale vegetation model relates the flood duration (i.e., the total number of flooding days per year), distance to the main channel, and land use (three types) to the presence/absence of

Table 1. Dominant biotypes in the Elbe floodplains

Biotype number	Biotype description
0	no data
1	Seasonally flooded grassland
2	Softwood floodplain forest
3	Hardwood floodplain forest
4	Reed
5	Herb fringes and herb meadows
6	Grassland of wet to moist sites
7	Intensively used, species-poor, moist grassland
8	Other reeds
9	Herby flood banks and -plains near the water
10	Dry and warm ruderal sites with dense vegetation
11	Moist ruderal sites

biotype groups. The model is a simplified version of the rules from MOVER (MOdel for VEgetation Response), which is a vegetation sub-model of INFORM (INtegrated FlOodplain Response Model) from the German Federal Institute (one of the main decision makers for the Elbe DSS) (Fuchs et al., 2002). In the MOVER model, a specification of the most expected biotype is determined on the basis of a set of rules. The rules consist of three matrices, one matrix for each land use type. The matrix gives for specific ranges of the distance to the main channel and ranges of the number of flooding days a biotype. Table 1 shows the dominant biotypes in the Elbe floodplains.

The flood duration can be determined from the critical discharge. The number of flooding days based on the critical discharge in the floodplain area is calculated for each cell (*x*,*y*) in the area using the approximation of the error function:

$$N_{flood}(x,y) = \frac{365}{2}\left(1 - erf\left[\frac{\log\left(Qcrit(x,y)\right) - \mu\left(x\right)}{\sigma\left(x\right)\sqrt{2}}\right]\right) \quad (4)$$

Where *Qcrit(x,y)* is the critical discharge; ¼$\left(x\right)$ and $\sigma(x)$ are mean value and standard deviation respectively. The frequency of biotypes is calculated as the individual number of cells of each biotype divided by the total number of cells of all 11 biotypes in the floodplains:

$$Fr_i = \frac{N_i}{N_{total}} \quad (5)$$

Where N_i is the number of cells of *i*th biotype in the floodplains; N_{total} is the total number of cells of biotypes in the floodplains; Fr_i is the frequency of *i*th biotype in the floodplains.

A detailed description of the models can be found in Fuchs et al. (2002).

Measures and Data

As shown in Figure 3, only two measures are considered to be relevant and implemented in this validation case study. One is groyne removing, i.e., 50% of the existing groynes will be removed from the main channel of the concerned river section. The effect of this measure is investigated by reducing Manning's roughness coefficients in the main channel. The other measure is renaturation of the floodplains. The effects of this measure are investigated by changing the land use types of meadow grass and agriculture in the left bank (floodplains) to broad-leaved forests. It is implemented in the HEC-6 model by mainly changing Manning's roughness coefficients in the floodplains. For the convenience of measure ranking, the current situation is represented as Measure 0 (M_0). Measure 1 (M_1) is groyne removing in the main channel and Measure 2 (M_2) is renaturation of the floodplains. Data used in this case study are provided by the German Federal Institute of Hydrology.

RESULTS FROM EACH STEP OF APPROPRIATENESS FRAMEWORK

Step 1 and Step 3: Uncertainty and Sensitivity Analysis

For the shipping model, three important sources of uncertainty are identified: (i) uncertainty in bed level measurements; (ii) uncertainty in discharge statistics; and (iii) uncertainty in the rating curves. For the vegetation model, only the uncertainties in the rating curves and discharge statistics are considered. Other sources of uncertainty are not taken into account mainly because of lack of information, although this does not mean that they are insignificant. Table 2 shows a brief introduction to the uncertainty in the above three sources.

Uncertainty in the Number of Shipping Days per Year for the Current Situation

For the current situation, the uncertainties in the aforementioned sources are propagated into the number of shipping days per year along the concerned river section by applying Latin Hypercube simulation (Saltelli et al., 2000). The simulation sample size is 1000. Figure 4 shows the error bar of the number of the shipping days per year for the current situation (M_0). The error bar contains three types of information: the 10 percentile of the model outputs, the mean value of the model outputs and the 90 percentile of the model outputs. Instead of standard deviations, the 10 and 90 percentiles are computed to get a better idea of the ranges of uncertainty in the model outputs. As shown here, the mean value of the number of shipping days per year is around 307 days. The results show that the uncertainty propagated into the number of shipping days is rather high (more than 10%).

Uncertainty in the Frequencies of Biotypes in the Floodplains for the Current Situation

In the vegetation model, uncertainty in the rating curves and discharge statistics are propagated into the model outputs. Figure 5 shows error bars of the frequencies of 11 biotypes in the floodplains along the concerned river section for the current situation. The 10 and 90 percentiles for the frequencies of all 11 biotypes are shown in this figure to get some indicators of the uncertainty in the frequencies of biotypes. This figure shows that the frequencies of biotype 2, 8, and 11 are quite small. The names of the biotypes can be found in Table 1. Biotype 0 represents the situation with no data available. The uncertainty in biotype 0 in Figure 5 therefore indicates variations of the

Table 2. Uncertainty sources

Uncertainty sources	Numbers	Inputs or parameters	Descriptions	Units	Uncertainty
Uncertainty in bed level	1	z	Bed level measurements	m	N(-, 0.10)
Uncertainty in discharge statistics	2	c_1	Regression parameter for discharge statistics μ at 340.0 -388.2 km	-	N(3.7×10^{-4},1.6×10^{-5})
	3	d_1	Regression parameter for discharge statistics μ at 340.0 -388.2 km	-	N(6.0,5.9×10^{-3})
	4	c_2	Regression parameter for discharge statistics σ at 340.0 -388.2 km	-	N(-6.3×10^{-5},1.2×10^{-5})
	5	d_2	Regression parameter for discharge statistics σ at 340.0 -388.2 km	-	N(6.0×10^{-1},4.3×10^{-3})
	6	c_3	Regression parameter for discharge statistics μ at 388.3 -422.6 km	-	N(9.0×10^{-5},2.7×10^{-5})
	7	d_3	Regression parameter for discharge statistics μ at 388.3 -422.6 km	-	N(6.1,1.1×10^{-2})
	8	c_4	Regression parameter for discharge statistics σ at 388.3 -422.6 km	-	N(-4.5×10^{-5},1.9×10^{-5})
	9	d_4	Regression parameter for discharge statistics σ at 388.3 -422.6 km	-	N(6.0×10^{-1},7.9×10^{-3})
	10	c_5	Regression parameter for discharge statistics μ at 422.7 -425.0 km	-	N(2.0×10^{-4},1.5×10^{-3})
	11	d_5	Regression parameter for discharge statistics μ at 422.7 -425.0 km	-	N(6.3,6.2×10^{-1})
	12	c_6	Regression parameter for discharge statistics σ at 422.7 -425.0 km	-	N(-5.0×10^{-5},1.1×10^{-3})
	13	d_6	Regression parameter for discharge statistics σ at 422.7 -425.0 km	-	N(5.9×10^{-1},4.5×10^{-1})
Uncertainty in the rating curves	14	e_1	Regression parameter for *a*	-	N(-1.5×10^{-1},1.2×10^{-3})
	15	f_1	Regression parameter for *a*	-	N(80,4.6×10^{-1})
	16	e_2	Regression parameter for *b*	-	N(2.2×10^{-4},2.8×10^{-3})
	17	f_2	Regression parameter for *b*	-	N(-2.4×10^{-2},2.8×10^{-3})

no-data situation due to the uncertainty of 11 biotype frequencies.

Sensitivity Analysis

Sensitivity analysis is performed to determine the order of importance of uncertainty sources in the

Figure 4. Error bars of the number of shipping days per year for the current situation and two measures (10 and 90 percentiles)

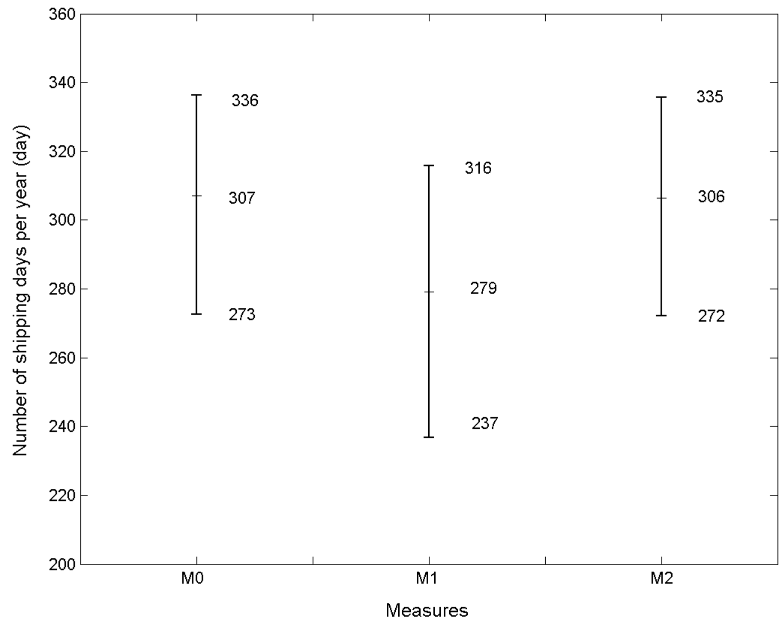


Figure 5. Error bars of the frequencies of 11 biotypes for the current situation (The error bar contains three types of information: the 10 percentile of the model outputs, the mean value of the model outputs and the 90 percentile of the model outputs)

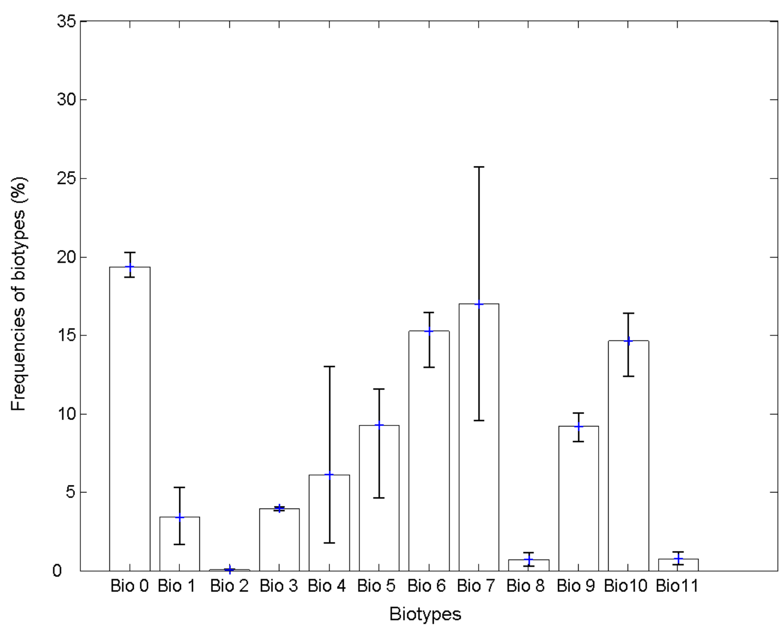
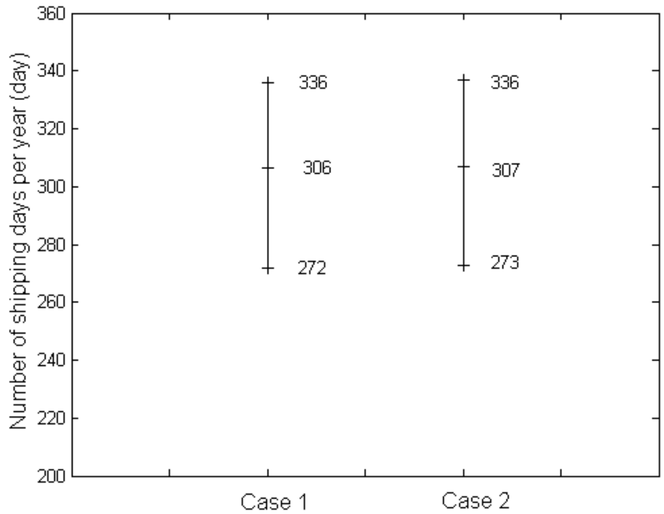


Figure 6. Error bars of the number of shipping days per year for two cases (Case 1 represents the situation with all uncertainties included; Case 2 represents the situation with only the uncertainty in the rating curves)



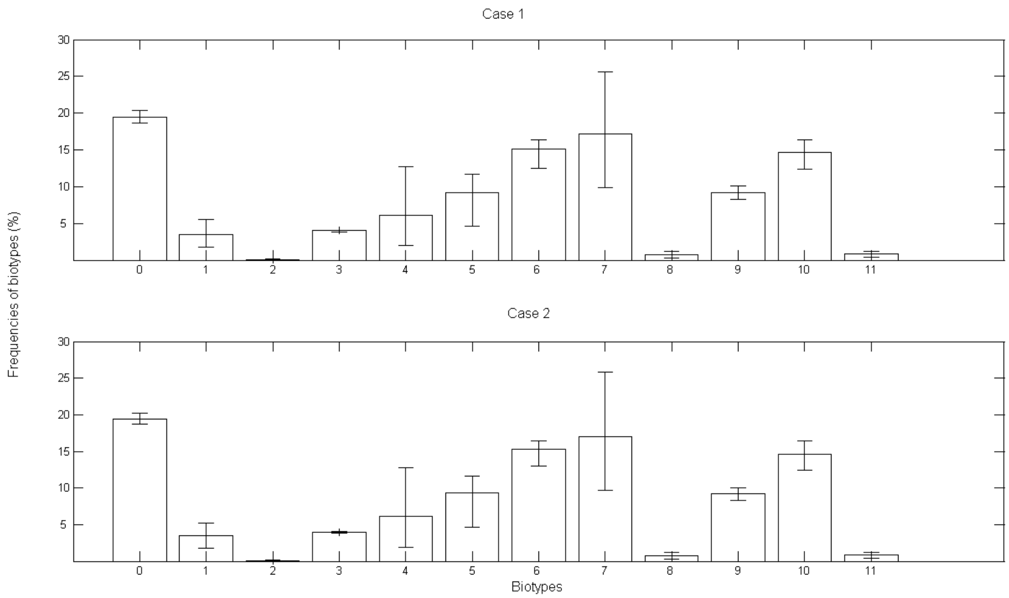
shipping and vegetation models. It is reasonable to regard that the uncertainty among three sources are independent. Therefore, an easy-going sensitivity analysis method is used by simply setting each source of uncertainty to zero, except those of interest. The uncertainty in the rating curves is the most important uncertainty source in the shipping model, followed by the uncertainty in bed level measurements and, lastly uncertainty in discharge statistics. The uncertainty in the rating curves is also the most important in the vegetation model followed by the uncertainty in discharge statistics. Thus for both models, the uncertainty in the rating curves is the dominant source of uncertainty resulting in a high uncertainty in the model outputs. Figures 6 and 7 show error bars of the number of shipping days per year and frequencies of biotypes, for the case with all uncertainties included and the case with only the uncertainty in the rating curves. It is shown in these two figures that the differences are insignificant for two cases.

Uncertainty Analysis for the Current Situation and Two Measures

Based on the results of sensitivity analysis, only the uncertainty in the rating curves is taken into account in the uncertainty analysis for the current situation and two different measures. Figure 4 shows error bars of the number of shipping days per year for these different situations.

It is reasonable to assume that the higher the number of shipping days per year, the better the measure is. If only mean values of the number of shipping days per year are considered, a ranking of measures can be easily obtained: $M_0 \succ M_2 \succ M_1$. Groyne removing (M_1) normally reduces water levels along the main channel, which may reduce the number of shipping days per year. Renaturation (M_2) supposes to increase water levels because of the increase of roughness in the floodplains and thus increases the number of shipping days per year. Here the existence of uncertainty in the model outputs gives a different ranking for M_0 and M_2. This not only means that the uncertainty in the model outputs is very high

Figure 7. Error bars of the frequencies of biotypes (Case 1 represents the situation with all uncertainties included; Case 2 represents the situation with only the uncertainty in the rating curves)



but also that renaturation has relatively small effects on the number of shipping days. From this point of view, the existence of high uncertainty in the model outputs determines that the ranking $M_0 \succ M_2 \succ M_1$ based on the mean values may not be accepted, particularly for M_0 and M_2.

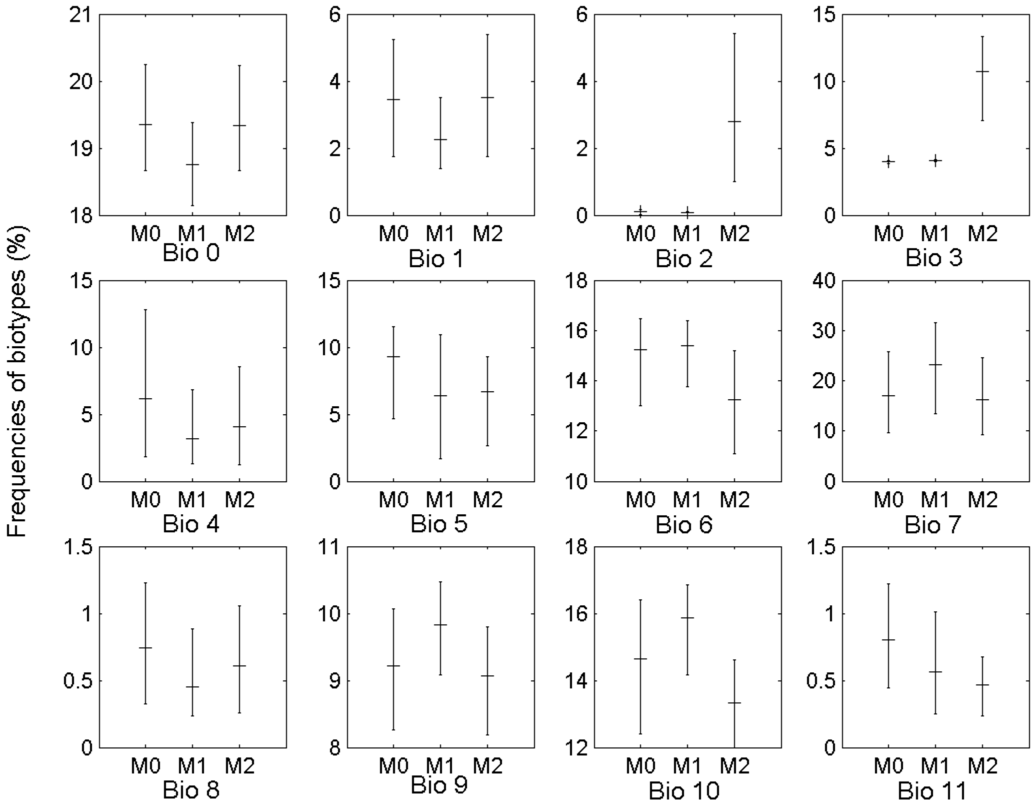
Figure 8 shows error bars of the frequencies of biotypes in the floodplains along the Elbe for the three different situations. As described before, M_2 is renaturation in the floodplains, which changes the land use types of meadow glass and agriculture into broad-leaved forest. M_2 has significant effects on the frequencies of biotype 2 and 3 instead of those of other biotypes, because biotype 2 and 3 are both wood (see Table 1). This can be observed from higher means in the error bars in Figure 5. Renaturation increased the number of cells (frequencies) of biotype 2 and 3 in the floodplains and decreased the number of cells of other biotypes. M_1 (groyne removing) has different effects on the frequencies of the 11 biotypes. If the measures aim to increase the frequencies of biotypes, a ranking of measures can be easily obtained if only the mean values are taken into account for each biotype. For example, for biotype 9, a ranking of measures can be obtained as $M_1 \succ M_0 \succ M_2$. However, the existence of high uncertainties in the frequencies of biotypes possibly makes this ranking unacceptable to decision-makers.

In order to determine whether or not the models used in the Elbe DSS are appropriate, the uncertainty analysis results need to be further analyzed. According to the appropriateness framework, the appropriateness of models is determined by calculating the risks of obtaining an unacceptable ranking for different pairs of measures. This will be dealt with in Step 2.

Step 2: Appropriateness Analysis

One recommendation described before is to consider the interdependency among model outputs from different measures. Reichert and Borsuk (2005) argued that if uncertainty in model outputs results from a common source of uncertainty that is

Figure 8. Error bars for the frequencies of biotypes for the current situation and two measures (10 and 90 percentiles)



not affected by measures, there is interdependency among model outputs from different measures. If the uncertainty source is affected by measures, there will be less interdependency among model outputs and therefore risks can be overestimated or underestimated. In the Elbe DSS case study, sensitivity analysis showed that the uncertainty from the rating curves is the dominant source and only this source of uncertainty has been considered in the uncertainty analysis of different measures for both shipping and vegetation models. As stated before, the two measures have been implemented in the HEC-6 model by changing the roughness of the main channel or the floodplains and thus changing the rating curves. This means that uncertainty in model outputs resulted from uncertainty in the rating curves, which are affected by measures. Therefore, it is reasonable to regard that uncertainty analysis will give independent model outputs for different measures for both shipping and vegetation models.

Questionnaire Results

In this validation case study, a questionnaire has been used to get a realistic value of the acceptable risk. It is assumed that the data from the questionnaire are representative for the views of two groups of participants. A decision analysis technique called Simple Additive Weighting is used to determine the acceptable risk (Malczewski, 1999). Equal weights are given to these two groups, representing equal importance. Based on the questionnaire results, for the shipping model,

Table 3. Mean differences of the number of shipping days per year and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures

Pair-wise measures	Mean difference (days)	Probability	Risk (days)
M_0 & M_1	28	0.24	7
M_0 & M_2	1	0.49	1
M_1 & M_2	27	0.24	7

the acceptable risk is estimated to be about 3 days (Xu, 2005). For the vegetation model, the acceptable risk is estimated to be around 6.5% of the frequencies of biotypes in the initial situation (Xu, 2005).

To obtain a desirable accuracy of questionnaire results, the number of participants is usually important. This number mainly depends on the proportion of participants responding to each of the categories in the questions, the precision required and the confidence level (Kish, 1995). If the questionnaire designs are simple random samples and the proportion is assumed to be 0.5, an approximate sample size needed for 5% precision with 95% confidence level is around 400 participants. The number of 30 participants in this case study can only result in a precision of 18%. Therefore, the number used in this questionnaire is possibly low in a statistical sense. However, in a real decision-making situation, to achieve a 5% precision, the number of decision-makers needs to be around 400 persons, which is seldom possible. Therefore, this study focuses more on whether or not these participants are representative. As stated before, the participants of the questionnaires are decision-makers and expert colleagues on modeling and uncertainty analysis. These people are regarded to be representative. Furthermore, the reliability of the questionnaires often depends on the reproducibility and high internal consistency of the questionnaire answers (Nadalet et al., 2005). Determining the reliability needs repeated questionnaires, which is not implemented in this case study.

Appropriateness of the Shipping Model

Table 3 shows the calculated mean differences of the number of shipping days per year and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures (M_0 & M_1, M_0 & M_2 and M_1 & M_2). As described before, the acceptable risk for the shipping model is around 3 days. In Table 3, the maximum risk of obtaining an unacceptable ranking computed is around 7 days, which is much higher than the acceptable risk. Therefore, it is concluded that the shipping model is not appropriate under the current uncertainty conditions.

Appropriateness of the Vegetation Model

As described before, the acceptable risk for the vegetation model is around 6.5%. However, for the easy understanding of the questionnaire, this value is a relative value which only indicates a proportion of the frequency of biotypes in the initial situation. It is therefore necessary to transform the relative acceptable risk to the absolute risk for 11 biotypes for the appropriateness analysis of the vegetation model (see Table 4).

Table 5 shows the calculated mean differences of the frequencies of 11 biotypes and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures. The risks are the most interesting parts of this table for the appropriateness analysis. The maximum risk for each biotype is compared with the corresponding

Table 4. Acceptable risks for 11 biotypes

Biotypes	Initial value (%)	Relative value of acceptable risk (%)	Absolute acceptable risk (%)
Biotype 0 (no data)	19	-	-
Biotype 1	3.5	6.5	0.23
Biotype 2	0.10	6.5	6.5×10^{-3}
Biotype 3	4.0	6.5	0.26
Biotype 4	6.1	6.5	0.40
Biotype 5	9.3	6.5	0.60
Biotype 6	15	6.5	0.98
Biotype 7	17	6.5	1.1
Biotype 8	0.74	6.5	4.8×10^{-2}
Biotype 9	9.2	6.5	0.60
Biotype 10	15	6.5	0.98
Biotype 11	0.81	6.5	5.3×10^{-2}

acceptable risk. Take biotype 1 as an example. The maximum risk comes from the pair of M_1 & M_2, with a value of 0.29%. This value is higher than the acceptable risk, which is 0.23%. Another example is biotype 9. The maximum risk comes from the pair of M_0 & M_1, with the value of 0.14%. Compare this value to the acceptable risk for biotype 9, which is 0.60%. The maximum risk is therefore smaller than the acceptable risk for biotype 9. Among the 11 biotypes, there are four biotypes for which the maximum risks are smaller than their corresponding acceptable risks. The other seven have risks that are larger than the acceptable risks (indicated by *) and are therefore regarded to be unacceptable.

If all frequencies of 11 biotypes in the floodplains are regarded as decision variables in the Elbe DSS, the vegetation model is determined to be inappropriate because the risks calculated for seven biotypes are unacceptable. However, if the decision variable is the frequency of only one biotype or some selected biotypes, for example biotype 10, the model can be determined to be appropriate because the risks of obtaining an unacceptable ranking are smaller than its acceptable risk. It is rational to regard that the higher the frequencies, the better the measures are. Then, the ranking for this biotype is $M_1 \succ M_0 \succ M_2$, which is determined based on the mean values of the frequency of biotype 10 in the floodplains. This ranking is regarded as acceptable because of the acceptable risks.

In this case study, it is assumed that all biotypes are equally important and they are all decision variables in the Elbe DSS. Therefore, the model is judged to be inappropriate because of the unacceptable risks for seven biotypes.

Step 4: Model Improvement

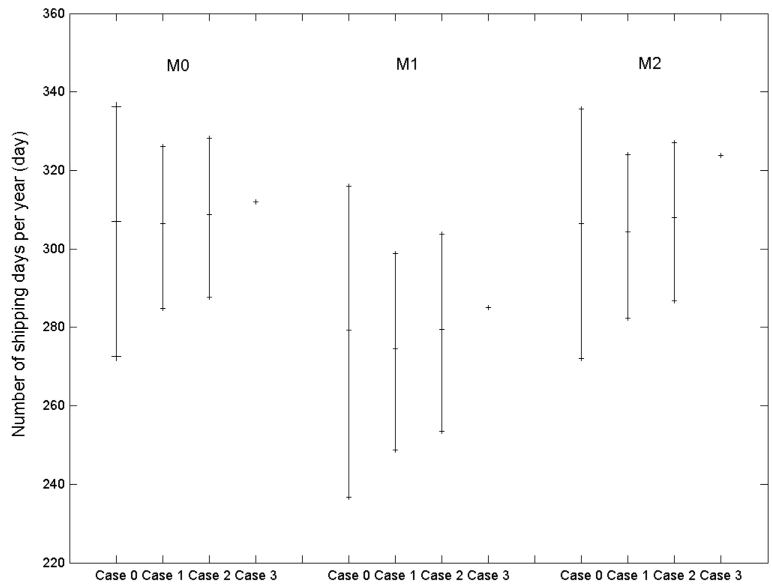
It has been concluded that both shipping and vegetation models are not appropriate due to the unacceptable high risks. According to the appropriateness framework, the models need to be improved after appropriateness analysis. As stated before, the uncertainty originating from the rating curves is the dominant source of uncertainty. Therefore, three improvements will be implemented in order to reduce the uncertainty from the rating curves based on the feasibility and availability of data and models. They are Case 1: use of more discharge inputs in the HEC-

Table 5. Calculated mean differences of the frequencies of 11 biotypes and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures

Biotypes	Pair-wise measures	Mean differences (%)	Probability	Risks (%)	Acceptable risk (%)
0 *No data*	M_0 & M_1	-	-	-	-
	M_0 & M_2	-	-	-	
	M_1 & M_2	-	-	-	
*1** *Seasonally flooded grass-land*	M_0 & M_1	1.2	0.25	0.29*	0.23
	M_0 & M_2	4.6×10^{-2}	0.48	2.2×10^{-2}	
	M_1 & M_2	1.2	0.24	0.29*	
*2** *Softwood floodplain forest*	M_0 & M_1	4.0×10^{-2}	0.29	1.1×10^{-2}*	6.5×10^{-3}
	M_0 & M_2	2.7	3.0×10^{-4}	8.1×10^{-4}	
	M_1 & M_2	2.7	0	0	
3 *Hardwood floodplain forest*	M_0 & M_1	6.3×10^{-2}	0.27	1.7×10^{-2}	0.26
	M_0 & M_2	6.7	0	0	
	M_1 & M_2	6.7	0	0	
*4** *Reed*	M_0 & M_1	2.9	0.25	0.74*	0.40
	M_0 & M_2	2.1	0.34	0.70*	
	M_1 & M_2	0.90	0.41	0.37	
*5** *Herb fringes and meadows*	M_0 & M_1	2.9	0.25	0.73*	0.60
	M_0 & M_2	2.6	0.22	0.58	
	M_1 & M_2	0.33	0.47	0.16	
6 *Grassland of wet to moist sites*	M_0 & M_1	0.15	0.49	7.2×10^{-2}	0.98
	M_0 & M_2	2.0	0.14	0.27	
	M_1 & M_2	2.1	0.11	0.24	
*7** *Intensively used, species-poor, moist grassland*	M_0 & M_1	6.2	0.25	1.5*	1.1
	M_0 & M_2	0.84	0.46	0.39	
	M_1 & M_2	7.1	0.22	1.5*	
*8** *Other reeds*	M_0 & M_1	0.30	0.24	7.0×10^{-2}*	4.8×10^{-2}
	M_0 & M_2	0.14	0.38	5.1×10^{-2}*	
	M_1 & M_2	0.16	0.34	5.4×10^{-2}*	
9 *Herby flood banks and -plains near the water*	M_0 & M_1	0.62	0.23	0.14	0.60
	M_0 & M_2	0.14	0.43	6.1×10^{-2}	
	M_1 & M_2	0.76	0.17	0.13	
10 *Dry and warm ruderal sites with dense vegetation*	M_0 & M_1	1.2	0.25	0.31	0.98
	M_0 & M_2	1.3	0.26	0.34	
	M_1 & M_2	2.6	6.5×10^{-2}	0.16	
*11** *Moist ruderal sites*	M_0 & M_1	0.24	0.27	6.5×10^{-2}*	5.3×10^{-2}
	M_0 & M_2	0.34	0.21	7.2×10^{-2}*	
	M_1 & M_2	0.10	0.43	4.2×10^{-2}	

*Calculated risk exceeds acceptable risk

Figure 9. Error bars of the number of shipping days per year for the current situation and two measures for four cases(10 and 90 percentiles)



6 model; Case 2: use of better calibration data; Case 3: improvement of the equation of the rating curves. The first two improvements are improvements of data quantity and quality and the third improvement is to improve the model equation. The original situation before the improvements is represented by Case 0.

More discharge data are used as inputs in the HEC-6 model to derive more discharge - water level data. The new rating curves at each location along the concerned river section are obtained based on these data (Case 1). Better calibration data have been obtained instead of the original calibration data (Case 2). The new model equation of the rating curves is:

$$H = H_0 + aQ^b \tag{6}$$

Where H is water level; H_0 is water-level correction factor; Q is discharge; a and b are location dependent parameters. This equation improves the original one (see Eq. (2)) because of the inclusion of the water level correction factor. This improvement is used to investigate the effect of model uncertainty on the appropriateness analysis of models (Case 3).

In the following, the effects of these improvements on the appropriateness analysis are investigated.

Effects on the Shipping Model

Figure 9 shows error bars of the number of shipping days per year for different measures for the original situation and the three improvements. The first two improvements (Case 1 and 2), being improving data quantity and quality, produced no significant differences between the mean value of the model outputs in the original situation (Case 0) and after the improvements. The mean of shipping days per year decreases slightly after improving data quantity and quality. The more important information in this figure is that the uncertainty ranges (the range between the 10 and 90 percentiles) become smaller, which indicates a

Table 6. Mean differences of the number of shipping days per year and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures for four cases

Pair-wise measures	Different cases	Mean differences (days)	Probability	Risks (days)
M_0 & M_1	Case 0	28	0.24	7
	Case 1	32	0.10	3
	Case 2	29	0.12	3
	Case 3	27	0	0
M_0 & M_2	Case 0	1	0.49	1
	Case 1	2	0.46	1
	Case 2	1	0.49	1
	Case 3	12	0	0
M_0 & M_2	Case 0	27	0.24	7
	Case 1	30	0.12	4
	Case 2	28	0.13	4
	Case 3	39	0	0

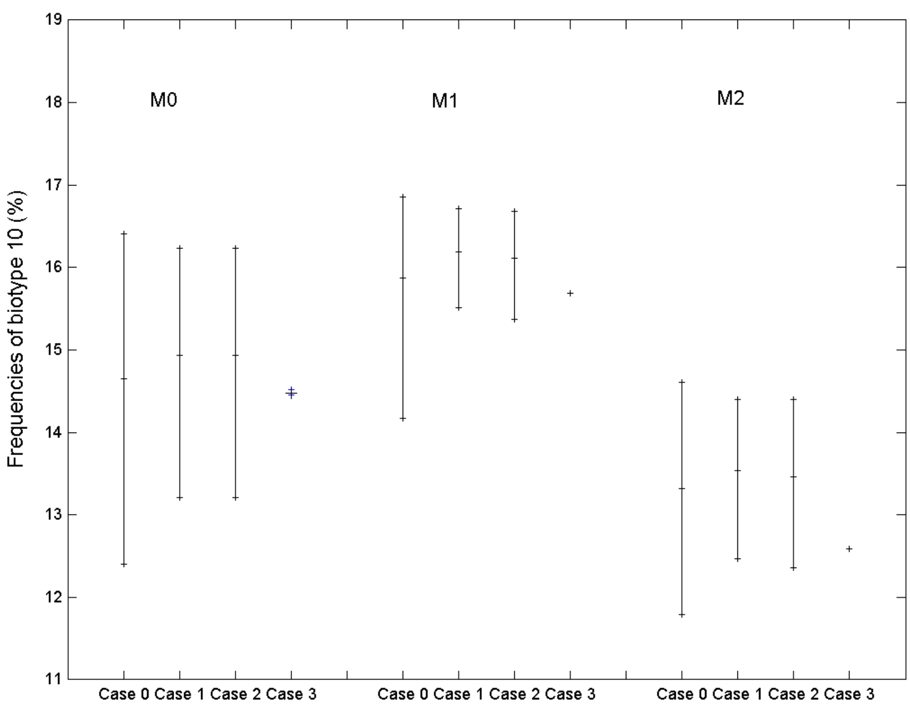
reduction of uncertainty in the number of shipping days per year. Improving the rating curve equation (Case 3), however, results in a small amount of uncertainty in the model outputs. This indicates a successful reduction of uncertainty. Adding a water level correction factor has improved the quality of the regression analysis greatly (see Eq. (2)). However, the nearly-deterministic model results may also indicate that other uncertainties are not considered because of lack of information while these uncertainties are possibly significant. For example, the model uncertainty from the shipping model structure is ignored. This can be improved in future by doing a thorough uncertainty analysis of the models if data are obtained.

Table 6 shows the mean differences of the model outputs and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures for the four cases. The maximum calculated risks of obtaining an unacceptable ranking for the first two improvements are smaller than those of the original situation, but still slightly higher than the acceptable risk of 3 days. This means that the shipping model is still inappropriate for decision-makers to make a comfortable ranking after improving data quantity and quality. For the third improvement, the calculated probabilities and risks are all zero, which indicates that the ranking can be made because the calculated risks are smaller than the acceptable risks. Therefore, the shipping model is regarded to be appropriate after this improvement.

Effects on the Vegetation Model

Figure 10 shows error bars for the frequencies of biotype 10 as an example under the original situation and the three improvements (four cases). For the first two improvements (Case 1 and Case 2), the lengths of the error bars are shorter than those in the original situation (Case 0). This indicates a reduction of uncertainty in the model outputs. The effect of improving of the rating curve equation (Case 3) is significant. The uncertainty in the frequency of biotype 10 has greatly deceased because of the improved equation. Therefore, the same series of discharge – water level data as in the original situation used for the improved equation showed a better fitting of the data to the rating curves and thus significantly reduced the

Figure 10. Error bars for the frequency of biotype 10 for the current situation and two measures for four cases (10 and 90 percentiles)



uncertainty in the rating curves. However, the significant reduction of uncertainty might also indicate that some other uncertainty sources are not considered here due to the availability of data and models.

Table 7 shows the mean differences of the model outputs and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures for the four cases from the vegetation model. As in the case of the shipping model, the effects of more discharge input and better calibration data on the appropriateness analysis results are small, still producing an unacceptable ranking of measures because of the unacceptable high risks. For the last improvement, due to small uncertainties in the model, the probabilities calculated are nearly zero for all 11 biotypes. The calculated risks of obtaining an unacceptable ranking are therefore also zero. In such a case, the risks are obviously smaller than the acceptable risks. The vegetation model is judged to be appropriate under the current uncertainty conditions.

CONCLUSION AND DISCUSSION

This paper validates the appropriateness framework using the Elbe DSS as a case study. The Elbe DSS case study started with inappropriate models. Three improvements have been used to reduce the uncertainty in model outputs to obtain an acceptable ranking of measures. After the third improvement of changing the model equation, models became appropriate. Therefore, the first validation criteria have been fulfilled. Furthermore, sensitivity analysis in the appropriateness framework identified the most important sources of uncertainty and the suggestions for model improvements were made accordingly. This is an efficient way to identify targets for reducing

Table 7. Mean differences of the frequency of biotype 10 and probabilities and risks of obtaining an unacceptable ranking for different pairs of measures for four cases

Pair-wise measures	Different cases	Mean differences (%)	Probability	Risks (%)
M_0 & M_1	Case 0	1.2	0.25	0.31
	Case 1	1.3	0.14	0.17
	Case 2	1.2	0.16	0.19
	Case 3	0	0	0
M_0 & M_2	Case 0	1.3	0.26	0.34
	Case 1	1.4	0.17	0.23
	Case 2	1.5	0.16	0.24
	Case 3	0	0	0
M_0 & M_2	Case 0	2.6	6.5×10^{-2}	0.16
	Case 1	2.7	6.5×10^{-3}	1.7×10^{-2}
	Case 2	2.6	9.1×10^{-3}	2.4×10^{-2}
	Case 3	0	0	0

the uncertainty. Although the final three improvements implemented are rather dependent on the availability of models and data, the last improvement derived appropriate models. Therefore, the appropriateness framework is regarded as efficient to achieve appropriate models and the second validation criterion has been satisfied as well. This shows that the appropriateness framework has been properly validated. Besides, the recommendations from the development case study were successfully implemented in the second validation case study. Therefore, this paper demonstrated that the appropriateness framework can be applied to other decision support systems in river basin management.

In this paper, the acceptable risk has been determined by sending questionnaires to relevant decision-makers and experts. A questionnaire by email has the advantages of ease, quick response, and if necessary, participation of a lot of people. However, some limitations about the questionnaire exist like the lack of face-to-face communication, which may lead to inaccuracy of questionnaire results. Therefore, in the Elbe DSS case study, several extra steps have been taken for a better understanding of questions, such as translation of the questionnaire to German and presentation to the participants. This hopefully can improve the quality of the questionnaire results. On the other hand, the number of participants who attends the questionnaires in the Elbe DSS case study is statistically low, illustrated by its low precision of 18% in the questionnaire results. To achieve a higher precision, the number of participants needs to be increased considerably. However, in reality, often the number of decision-makers involved is not big enough to achieve this high precision. One solution is to ask relevant experts to participate in the questionnaire and verify the results, as has been done in this paper.

The analysis done in this paper is highly dependent on the available models and data provided by the Elbe DSS Project. The improvements to reduce the uncertainty in the data and models are mainly based on the identified uncertainty sources, and the dominant sources of uncertainty have been investigated to improve the models. If more information on unknown and/or unidentified uncertainty sources in the models can be obtained, a thorough uncertainty analysis is highly recommended and needs to be combined into the appropriateness analysis. Furthermore, because of the availability of models and data, the ways to reduce the uncertainty are rather limited. Possible improvements of hydraulic models, for example 2-D hydrodynamic models, are not considered. Therefore, an investigation of the effect of more complex hydraulic models on the appropriateness analysis (or ranking of measures) could not be implemented. This can be investigated in a future study if models are available. However, it should be kept in mind that more complex model may introduce more uncertainty.

The usefulness of the appropriateness framework lies in adding more reliability to decision-making under uncertainty, stimulating better applications of models in decision-making, avoiding the development of over-complex models and promoting better communication between modelers and decision-makers. The current appropriateness framework is suggested to be applied in situations where decision-makers and modelers interact in the context of problem solving. Decision-makers need to pose their problems and management objectives clearly to modelers or analysts. The latter can then take this into account when making or choosing models that can solve decision-makers' problems and reach their objectives. Close interactions and dialogues between decision-makers and modelers are therefore substantial to set up an appropriate model based on management objectives and help decision-makers to tackle the problem of uncertainty.

The appropriateness framework proposed is based on several assumptions. First, decisions are supposed to be made on a rational basis. Second, decision variables considered in the DSS should be quantifiable. Finally, it should be possible to obtain a reasonable estimation of the levels of uncertainty involved in decision-making. However, one has to realize that in reality decisions are not always made on a rational and quantitative basis. For example, Van Asselt (2000) argued that there is political uncertainty, which may complicate decision-making. In some cases, certain disasters may play an important role in decision-making as well. One example is that the Dutch government only started to regard the flooding problem as serious after the flooding in the Meuse River Basin in 1993 and 1995. These assumptions should be kept in mind when applying the appropriateness framework.

ACKNOWLEDGMENT

This work was part of a research project funded by both the Royal Netherlands Academy of Arts and Sciences and the University of Twente. The authors would like to thank the German Federal Institute of Hydrology for providing the data for the case study used in this paper. The authors are also grateful for valuable comments from Prof. Arthur E. Mynett from WL| Delft Hydraulics on this work and for the help from Dr. Jean-Luc De Kok from the University of Twente in preparing the data and questionnaires. Finally, the authors would like to thank Prof. Yeou-Koung Tung from Hong Kong University of Science and Technology for providing valuable comments on the final completion of this paper.

REFERENCES

Bevington, P. R., & Robinson, D. K. (1992). *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill.

Booij, M. J. (2005). Impact of climate change on river flooding assessed with different spatial model resolutions. *Journal of Hydrology (Amsterdam)*, *303*(1-4), 176–198. doi:10.1016/j.jhydrol.2004.07.013

Boorman, D. B., Williams, R. J., Hutchins, M. G., Penning, E., Groot, S., & Icke, J. (2007). A model selection protocol to support the use of models for water management. *Hydrology and Earth System Sciences*, *11*, 634–646. doi:10.5194/hess-11-634-2007

Brooks, R. J., & Tobias, A. M. (1996). Choosing the best model: level of detail, complexity, and model performance. *Mathematical and Computer Modelling*, *24*(4), 1–14. doi:10.1016/0895-7177(96)00103-3

Busemeyer, J. R., & Wang, Y. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*(1), 171–189. doi:10.1006/jmps.1999.1282

De Kok, J. L., Wind, H. G., Delden, H., & Verbeek, M. (2000). *Towards a Generic Tool for River Basin Management. Feasibility Assessment for a Prototype DSS for the Elbe. Feasibility Study – report 2/3. Final report*. Enschede, the Netherlands: University of Twente.

Devooght, J. (1998). Model uncertainty and model inaccuracy. *Reliability Engineering & System Safety*, *59*, 171–185. doi:10.1016/S0951-8320(97)00137-3

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B. Methodological*, *57*(1), 45–97.

Forster, M. R. (2000). Key concepts in model selection: performance and generalizability. *Journal of Mathematical Psychology*, *44*, 205–231. doi:10.1006/jmps.1999.1284

Fuchs, E., Giebel, H., Hettrich, A., Huesing, V., Rosenzweig, S., & Theis, H. J. (2002). Einzatz von ökologischen Modellen in der Wasser- und Schiffahrtsverwaltung, das integrierte Flussauenmodell INFORM (Applications of Ecological Models in Water and Navigation Management, the Integrated Model INFORM). BfG Mitteilung Nr. 25, Koblenz.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics Series*. New York: Chapman and Hall.

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, *29*(8), 2637–2649. doi:10.1029/93WR00877

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*, 602–614. doi:10.1016/j.envsoft.2006.01.004

Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, *1*(1), 11–27. doi:10.1111/j.1539-6924.1981.tb01350.x

Kish, L. (1995). *Survey Sampling*. New York: Wiley-Interscience.

Krysanova, V., Hattermann, F., & Wechsung, F. (2004). Implications of complexity and uncertainty for integrated modelling and impact assessment in river basins. In: Pahl-Wostl, C., Schmidt, S., Rizzoli, A. E., & Jakeman, A. J. (Ed.), *Complexity and Integrated Resources Management, Transactions of the 2nd Biennial Meeting of the International Environmental Modeling and Software Society.* iEMSs: Manno, Switzerland.

Loague, K. M., & Freeze, R. A. (1985). A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research*, *21*(2), 229–248. doi:10.1029/WR021i002p00229

Malczewski, J. (1999). GIS and Multicriteria Decision Analysis. New York: John Wiley & Sons, INC.

Nadalet, L., Kohl, F., Pringuey, D., & Berthier, F. (2005). Validation of a subjective quality of life questionnaire in schizophrenia. *Schizophrenia Research*, *76*, 73–81. doi:10.1016/j.schres.2004.09.017

Neuman, S. P. (2004). Maximum likelihood Bayesian averaging of uncertain model predictions. [SERRA]. *Stochastic Environmental Research and Risk Assessment*, *17*(5), 291–305. doi:10.1007/s00477-003-0151-7

Palmer, R. N., & Cohan, J. L. (1986). Complexity in Columbia River Systems Modeling. *Journal of Water Resources Planning and Management*, *112*(4), 453–468. doi:10.1061/(ASCE)0733-9496(1986)112:4(453)

Parker, M., Thompson, J. G., Reynolds, R. R., & Smith, M. D. (1995). Use and misuse of complex models: examples from water demand management. *Water Resources Bulletin*, *31*, 257–263.

Reichert, P., & Borsuk, M. E. (2005). Does high forecast uncertainty preclude effective decision support? *Environmental Modelling & Software*, *20*(8), 991–1001. doi:10.1016/j.envsoft.2004.10.005

Robinson, J. B., & Freebairn, D. M. (2001, December 10-13). The decision- maker has an important, but often neglected, role in model use. In: Proceedings of MODSIM 2001, *International Congress on Modelling and Simulation*. Canberra, Australia.

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity Analysis*. New York: John Wiley & Sons Ltd.

Shaw, E. M. (1994). *Hydrology in Practice*. Cornwall, UK: T.J. International Ltd, Padstow.

Turner, D. P., Dodson, R., & Marks, D. (1996). Comparison of alternative spatial resolutions in the application of a spatially distributed biogeochemical model over complex terrain. *Ecological Modelling*, *90*(1), 53–67. doi:10.1016/0304-3800(95)00143-3

U.S. Army Corps of Engineers. (1993). *CPD-6, HEC-6, Scour and Deposition in Rivers and Reservoirs*. User's Manual.

Van Asselt, M. B. A. (2000). *Perspectives on Uncertainty and Risk. The PRIMA Approach to Decision. Support*. Amsterdam: Kluwer Academic Publishers.

Vreugdenhil, C. B. (2002). Accuracy and reliability of numerical river models. *Journal of the American Water Resources Association*, *38*(4), 1083–1095. doi:10.1111/j.1752-1688.2002.tb05548.x

Vreugdenhil, C. B. (2006). Appropriate models and uncertainties. *Coastal Engineering*, *53*, 303–310. doi:10.1016/j.coastaleng.2005.10.017

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*, 92–107. doi:10.1006/jmps.1999.1278

Xu, Y. P. (2005). *Appropriate modeling in decision support system for river basin management.* (PhD thesis), University of Twente, Enschede, the Netherlands.

Xu, Y. P., Booij, M. J., & Mynett, A. (2007). An Appropriateness Framework for the Dutch Meuse Decision Support System. *Environmental Modelling & Software*, *22*(11), 1667–1678. doi:10.1016/j.envsoft.2007.01.002