Bayesian Networks and Social Objectives: A Canoeing Case Study

Ed Shaw

Vikas Kumar

D. N. Lerner

Eckart Lange

International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Bayesian Networks and Social Objectives: A Canoeing Case Study

Ed Shaw[1*], Vikas Kumar[1], David Lerner[1] & Eckart Lange[2]
[1] *Department of Civil and Structural Engineering, Mappin Street, Sheffield, S1 3JD, UK*
[2] *Department of Landscape, Crookesmoor Building, Conduit Road, Sheffield, S10 1FL, UK*
[*] *email:ed.shaw@sheffield.ac.uk*

Abstract: Much of the value of a river lies in the sociocultural values people attach to it. Though such 'social objectives' should form a fundamental part of integrated catchment management, they have typically been neglected due to the qualitative nature of many of the variables that predict them. One such social objective in the Don catchment, UK, is the management goal of maximising the recreational quality of rivers for canoeing within the constraints imposed by other management aims such as reducing flooding. Recreational quality is impacted by the modification of river weirs, an important management intervention in the catchment for which there are multiple potential options. An integrated catchment management decision support tool must predict the impact of the different modification options, raising the question; how to deal with the complex, uncertain, and subjective variables that determine a canoeist's judgement of river quality? To tackle this issue, we employed a Bayesian Network (BN), which uses probability to describe relationships between variables. As probability can incorporate expert estimates, this enabled us to harness the knowledge of canoeist stakeholders. In this paper we discuss the experiences of building a BN with the collaboration of canoeists to predict how weir modification affects river recreational quality, and comment on implications for the overall utility of the approach for modelling social objectives. We conclude BNs are indeed suitable for modelling social objectives; probabilities capturing the uncertainty and subjectivity of variables such as 'weir danger'. However the approach also has clear limitations. Though the canoeists found most parts of BN construction engaging, this was not so when eliciting judgements of probability, for which the use of questionnaires made it an abstract process. A further issue was that the number of questions needed in the probability elicitation stage increased dramatically with BN complexity. Interpolating probabilities from a limited number of questions is a partial solution, but even so the high number of questions still necessary was a barrier to completing this step. Ultimately this will constrain the potential complexity of BNs that require expert knowledge to define probabilistic relationships.

Keywords: Integrated catchment management; Bayesian Network; decision support; social objectives; canoeing

1. INTRODUCTION

A large component of the value of a river lies in the sociocultural values people attach to it. These include aesthetic, recreational, cultural and spiritual perceptions of value in the existence of, or in an individual's experience with, elements or states of an ecosystem. Collectively these values are known as non-use or sociocultural ecosystem services [Millennium Ecosystem Assessment, 2005]. Predicting how river interventions affect these values must form a fundamental part of integrated decision support models built to support urban river management. However, such 'social objectives' have tended to be neglected from process based models due to the qualitative nature of many of the variables that need to be included within the model structure [Holzkaemper et al., submitted]. In addition water management policy has traditionally focussed on physical objectives such as chemical water quality meaning there has been a lack of a reason why social objectives should be considered in the decision making process, though this is currently changing with the introduction of the Water Framework Directive in the European Union which provides some scope for social objectives to be traded-off against environmental objectives. The integration of human perceptions and values with catchment management models is now considered a key research direction in developing model based tools to aid water management [Borowski & Hare, 2007]. Yet if this integration of social objectives is to be achieved, then the problems that come with it; complexity, uncertainty, and subjectivity of the relationships between variables, must be overcome.

One solution is to take the conceptual modelling approach of a Bayesian Network (BN). In the last decade BNs have increasingly been applied to environmental management problems, and recently also to integrated water management issues [Ames et al., 2005; Barton et al., 2008; Kumar et al., 2008]. BNs use a graphical cause-effect network using probabilities to describe the conditional relationships between the dependent variables (known as child nodes) and their controlling variables (parent nodes). Each variable is described as a range of discrete states so that the model user can specify the state of parent nodes from which the BN predicts the likelihood that each child node state will result. Child nodes may also be parent nodes of further child nodes creating a chain so that the probabilities propagate through the system until a basal child node is reached. By using probabilities, relationships between variables can be derived from expert judgement, capturing uncertainties in expert knowledge.

These qualities make a BN approach suitable for modelling the relationships between weir modification and the recreational quality of rivers for canoeing. The Graphical User Interface (GUI) of a cause-effect network means the model is relatively transparent, with assumed relationships between variables clear to see. By describing relationships between variables probabilistically, mentally held perceptions and judgements of value can be described, allowing the views of stakeholders to be harnessed. Also, probabilities can capture the uncertain nature inherent in subjective variables, where potentially every individual can hold a unique view.

In this paper we discuss the experiences of building a BN that addresses a high profile management problem in the Don Catchment, UK; that of the impact of weir modification on the quality of the River Don for canoeing. We use the experiences of building the BN with the collaboration of canoeing groups to explore issues regarding the utility of the approach for modelling social objectives. What did and didn’t work well in the process of building the model is examined and the suitability of BNs is commented on.

2. CASE STUDY

A current major management problem facing decision makers in the Don Catchment, UK, and in many other catchments around the world is that of the modification of the many weirs that impound river systems to mitigate their negative environmental impacts. The Don Catchment has a particularly high number of weirs as it was a historically important centre of water powered metal working. Despite most now being obsolete, weirs are still a common sight in the Don Catchment (see Figure 1). Many groups would like to modify the weirs, with some anglers wanting to install fish passes; potential victims of flooding interested in removing weirs; proponents of hydro-electricity wishing to install micro-hydro schemes; enthusiasts of natural history aspiring to restore the river to a more natural state and enthusiasts of the river's heritage wanting to conserve weirs for their historic value.

All these different modifications have potential to affect the recreational quality of the rivers for another stakeholder group; that of the local canoeists. Canoeing is a popular leisure activity in the city of Sheffield through which the River Don flows. The various weir modification options have both positive and negative effects on the recreational quality of this activity, though the impact is subjective; based on the judgement of the local canoeing groups. If weir modification decisions are to be aided by an integrated decision support tool, then predicting this impact on recreational quality must be included as a social objective.

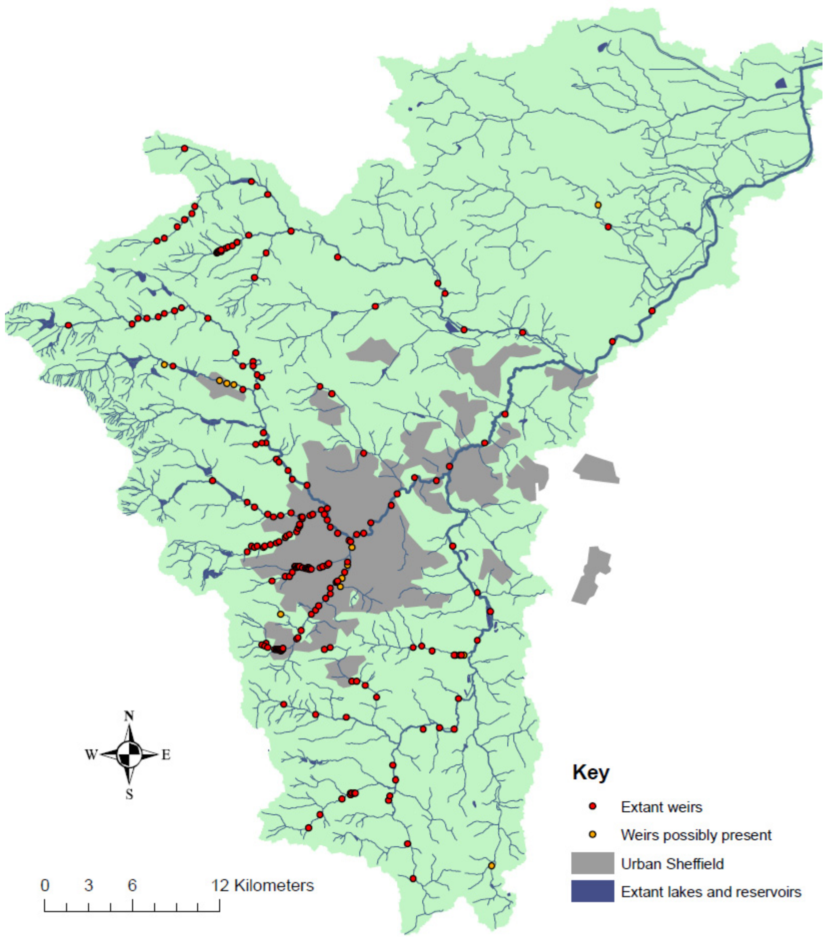


Figure 1. The distribution of weirs in the Don Catchment.

3. THE CONSTRUCTION OF THE CANOEING BN

In the following sections we describe the four steps to the process of constructing the BN built to predict the impact of weir modification on the recreational quality of rivers. This process of deriving expert judgement to construct the model is known as knowledge elicitation.

3.1 Identification of model variables and structure

The first step of building a BN is to identify the variables and structure of the cause-effect network. Two workshops were held to achieve this step, with canoeists from local groups invited to take part. Six local canoeing groups were identified using the internet or through contacts and invited to attend, of which three agreed to send representatives providing a total of five participants. It is not clear why three groups didn't participate but for these we only had email contact details and it might reflect the impersonality, and frequency of junk mail associated with this communication medium. The process of creating the structure started as an informal group brainstorming exercise, where the canoeists were encouraged to think about how a weir could affect the quality of a river for canoeing, a variable that was to form the basal child node in the network. The canoeists identified that a weir could affect quality in two ways; as a source of danger, and as a source of fun (descending weirs was considered exciting), so these became parent nodes to the child node. The process was repeated, this time by thinking about what factors determined the danger and fun of the weir, and so on until measurable quantitative variables were identified that could form the input nodes to the structure. The next step was to factor in the impact of weir modification options (changing weir height, steepness, orientation, profile of weir face, and installation of microhydro, canoe pass and fish passes), so these were described to the canoeists who then discussed as to whether they would influence any of the model's variables. Lastly, the nodes least important to the model were examined and dropped if it was thought that they wouldn't make much difference to the output of their child nodes (see Figure 2 for an overview of the evolution of the network).

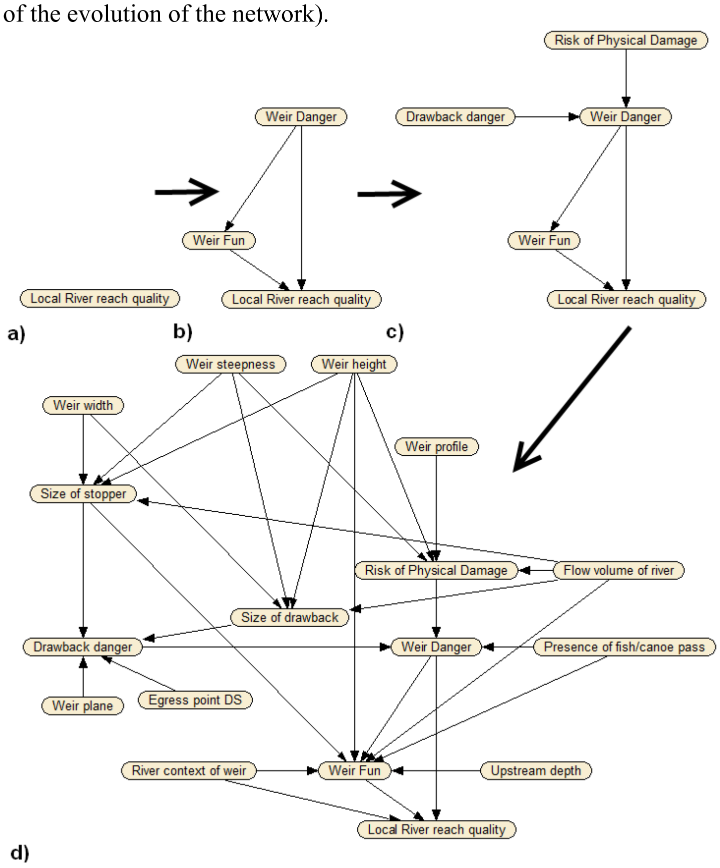


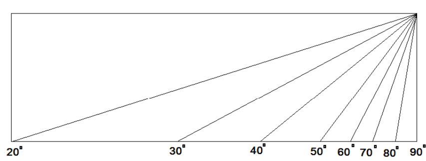
Figure 2. The evolution of the BN structure in the identification of model variables and structure stage. **a)** The basal child node of river quality, **b)** Weir fun and weir danger were identified as the two main ways weirs impact on river quality for canoeing, **c)** Weir danger was found to be controlled by the weir drawback (cyclical river flow at the base of the weir) and risk of physical injury descending the weir, **d)** The final canoeing BN structure with all remaining parent nodes and linkages identified.

We consider this stage of model construction to have worked very well. Those involved found the cause-effect network intuitive, and seemed very engaged; enjoying talking about their hobby. The finished cause effect network could then be directly entered into the BN.

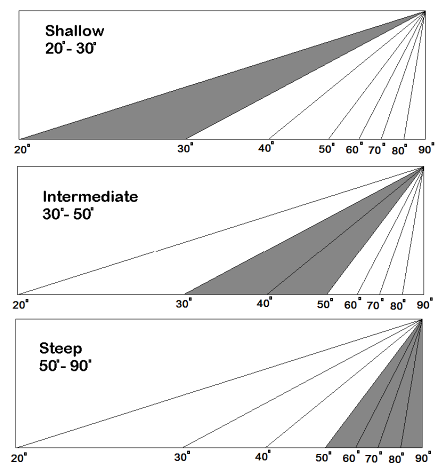
3.2 Categorisation of the variables

One of the characteristics of BNs is that model variables are described as discrete states e.g. weir height can be described as tall, medium and small. This means that the variables have to be defined so that they mean the same thing to all involved in the use and construction of the model, and so that the threshold between states is at a point that results in a meaningful change in the predicted states of the dependent variable. For example, changing weir height from 2m to 3m tall makes a big difference to the risk a canoeist faces descending a weir, but a change in height of 10m to 11m makes little difference.

This process was also conducted at the same workshops and was aided using prepared visual aids for those variables anticipated likely to be important to the canoeists (e.g. see Figure 3). Once again, the canoeists completed this process with relative ease and seemed to enjoy the process.



a)



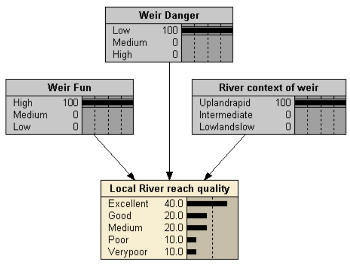
b)

Figure 3. a) Visual aid used to help the canoeists classify the states for the variable weir steepness. **b)** The resulting ranges of weir steepness allocated to the discrete states of shallow, intermediate and steep.

3.3 Probability elicitation

In contrast to the previous steps, the process of eliciting the probabilities caused problems for the model construction. This stage requires the completion of probability tables for each relationship between a child and its parent nodes (e.g. see Figure 4), where the expert estimates the likelihood of child node being in each state for every permutation of parent node states. However, as the complexity of the model increases, then the number of permutations increases exponentially, so that for example, the sub-network of weir fun with its seven parents has 2916 potential combination of states. This would require the expert to answer 2916 questions to fill out the conditional probability table for this sub-network alone! Clearly this is not acceptable, and for this reason we used an approach developed by Kumar et al. (in prep.) which only requires the expert to provide estimates of probabilities for a limited number of parent node states, from which the remaining probabilities are interpolated. Using this method we reduced the number of questions down to 120.

After the workshops, all participants agreed to complete the questionnaires, which had to be constructed using the outputs of the workshops. Of the five copies of the questionnaire sent to participants, and despite repeated reminders and offers of help, not a single copy was returned. As this first attempt at probability elicitation had failed, a change in tactic was tried. Supposing that that the number of questions on the questionnaire was still unreasonably long, and the activity was asocial as compared to the workshops the participants had enjoyed, new canoeing contacts were made and were personally supervised when filling out the questionnaires. The process of completing the questionnaires was very time consuming, taking between 2 to 5 hrs. Participants needed to spend time thinking about the questions and often referred to the supervising helper for further explanation, and to check that they had interpreted the questions correctly.



a)

Weir Danger	Weir Fun	River context of weir	Excellent	Good	Medium	Poor	Verypoor
Low	High	Uplandrapid	40.000	20.000	20.000	10.000	10.000
Low	High	Intermediate	27.200	20.000	32.800	10.000	10.000
Low	High	Lowlandslow	27.200	13.600	16.800	13.200	29.200
Low	Medium	Uplandrapid	28.800	20.000	31.200	10.000	10.000
Low	Medium	Intermediate	16.000	20.000	44.000	10.000	10.000
Low	Medium	Lowlandslow	16.000	13.600	28.000	13.200	29.200
Low	Low	Uplandrapid	28.800	14.400	17.200	12.800	26.800
Low	Low	Intermediate	16.000	14.400	30.000	12.800	26.800
Low	Low	Lowlandslow	16.000	8.000	14.000	16.000	46.000
Medium	High	Uplandrapid	24.000	20.000	36.000	10.000	10.000
Medium	High	Intermediate	11.200	20.000	48.800	10.000	10.000
Medium	High	Lowlandslow	11.200	13.600	32.800	13.200	29.200
Medium	Medium	Uplandrapid	12.800	20.000	47.200	10.000	10.000
Medium	Medium	Intermediate	0.000	20.000	60.000	10.000	10.000
Medium	Medium	Lowlandslow	0.000	13.600	44.000	13.200	29.200
Medium	Low	Uplandrapid	12.800	14.400	33.200	12.800	26.800
Medium	Low	Intermediate	0.000	14.400	46.000	12.800	26.800
Medium	Low	Lowlandslow	0.000	8.000	30.000	16.000	46.000
High	High	Uplandrapid	24.000	12.000	16.000	14.000	34.000
High	High	Intermediate	11.200	12.000	28.800	14.000	34.000
High	High	Lowlandslow	11.200	5.600	12.800	17.200	53.200
High	Medium	Uplandrapid	12.800	12.000	27.200	14.000	34.000
High	Medium	Intermediate	0.000	12.000	40.000	14.000	34.000
High	Medium	Lowlandslow	0.000	5.600	24.000	17.200	53.200
High	Low	Uplandrapid	12.800	6.400	13.200	16.800	50.800
High	Low	Intermediate	0.000	6.400	26.000	16.800	50.800
High	Low	Lowlandslow	0.000	0.000	10.000	20.000	70.000

b)

Figure 4. a) Subnetwork of canoeing BN and **b)** corresponding conditional probability table

3.4 Model validation

At the point of writing, the model still needs to be validated. There are two planned parts to this stage, both requiring the participation of canoeists. Firstly, the canoeists are to be introduced to the canoeing BN compiled using the data derived from their questionnaires. They will be encouraged to play with the model, testing different permutations of weir states, and to examine if the predicted child node states meet their expectations. New values will be entered into the model if they don't feel that the predicted outcomes are satisfactory. Once this refinement has occurred, the canoeists will be invited to canoe a length of the River Don in Sheffield. For each weir descended, they will each rate its characteristics e.g. weir fun and danger, and these compared to the weir fun and danger states predicted by the BN. Involvement of canoeists who didn't participate in model construction will also be sought in order to test how reflective of the model is of the wider canoeing community. We believe this should provide a robust assessment of the canoeing BN's validity.

4. DISCUSSION AND CONCLUSION

In creating a BN to model the impact of weir modification on the quality of the River Don for canoeing, we believe it has been shown that this technique is appropriate for modelling social objectives. With a working model and clear path to validation and improvement, it is expected that the predictive ability of the final version will be fairly reliable. By using conditional probabilities to describe relationships between variables, the approach successfully captured subjective concepts such as 'weir danger' or 'weir fun', and the inherent uncertainty in the relationships between these qualitative variables.

As stakeholders must provide the knowledge describing the relationships between variables when modelling social objectives, then it is vital that the process of constructing the model must be stakeholder friendly. In this regard the BN was generally successful. The canoeists found the steps of identifying the variables and structure of the model, and categorising the variables as engaging and relatively intuitive processes. However as in the experience of

Henriksen et al. [2007] who found aspects of BNs difficult for non-experts to understand, the canoeists also struggled with one part of creating the BN. In our case this was linked to the probability elicitation stage, required to create the conditional probability tables. While we are not certain why no questionnaires were initially returned, it seems reasonable to assume that it was due to the combination of the length and the abstract nature of the questionnaire. Trying to envisage multiple states of a set of parent nodes described textually is difficult, and indeed those supervised to fill out the questionnaire found it challenging. The large number of questions required in the questionnaire, an outcome of dependence on expert knowledge, was also demanding of the stakeholders. As the number of variables in a BN increases linearly, the number of questions needed to fill out the conditional probability tables increases exponentially. Consequently the canoeing BN would not have been practical to build without the method of Kumar et al., [in prep.] to interpolate probabilities from a limited number of questions. Even so the number still required was large and a barrier to completing the knowledge elicitation stage. For some BNs relationships between physical variables can be described using empirical data or data from other pre-existing models, but neither were available for any of the relationships in the canoeing BN. In hindsight more effort should have been taken to simplify the model by removing less important variables, though there would have been an erosion of model usefulness. Ultimately the excessive demands put on experts by questionnaires will constrain the potential complexity of BNs requiring expert knowledge to derive conditional probabilities. Methods need to be developed to increase the ease of completing the probability elicitation process to reduce the burden on those providing the expert knowledge.

Post-construction BNs are also relatively user friendly for stakeholders; important as involvement of stakeholders in the decision making process is considered important if decisions are to be fair and sustainable [Soncini-Sessa, 2007]. With a Graphical User Interface (GUI) of an interactive cause-effect network, BNs are intuitive way to explore how a weir modification options affect river quality for canoeists, and this transparency builds trust. By displaying the likelihood that certain states will be realised due to weir modification, then the inherent uncertainty in the model is communicated to the user (though BNs don't differentiate between uncertainty in the system and in stakeholder understanding). As the BN structure and variables are defined by stakeholders, then it is automatically constructed at a level appropriate to the stakeholders and with relevant indicators defined. This user friendliness makes it an inclusive tool, meaning BNs can be used to promote understanding between various stakeholder groups, as well as providing information to support decision making. Even so, there is still some scope for the improvement of the GUI. Consider the canoeing BN with its ten input variables. Mentally combining these attributes and visualising actual weir and river they represent is difficult. Given that water managers do not have time to learn or teach the public how to use models that can be difficult to understand [Borowski et al., 2007] then the reliance on abstract mental visualisation is probably a barrier to BN uptake. To improve the user friendliness of the GUI is one reason why Gill et al., [2010] have linked the canoeing BN introduced in this paper to interactive visualisation software. The visualisation communicates what a weir based on user selected input BN states would look like, and allows the user to interact with the weir e.g. changing height, designing different weir modification scenarios, making the BN more accessible to participants in decision making.

However, in addition to the knowledge elicitation difficulties at the probability elicitation stage, there were some other drawbacks in using the BN approach for modelling the canoeing social objective. BNs cannot easily deal with spatial information, meaning that while the canoeing BN could predict how a single stretch of river quality was affected by weir modification, it couldn't consider the cumulative effect of multiple weirs on longer stretches of rivers. Other approaches must be used in conjunction with BNs if spatial issues are to be dealt with. BNs may also have a limitation in their function as tools for decision makers. Borowski et al., [2007] found there was a need for models that helped the user to develop new solutions to management problems. As discrete management options are predefined in the BN then scope for users to design new management options is restricted. This limitation can be partially overcome by allowing and encouraging users to modify the BN, incorporating new ideas after its initial construction. If the stakeholders don't know what the effects of a new modification variable will be, then this means that it cannot be included until this understanding is gained. A further issue is that the categorisation of the variables into discrete states makes the BN unresponsive to small changes in input variables, meaning the tool is not useful for fine tuning designs.

Lastly, some questions still remain on applicability of the BN approach for modelling of social objectives. While the relationships and variables involved in determining the effect of weir modification on the quality of the river for canoeing were clear cut, this may not be the case for other social objectives. Indeed, some cultural values such as perceptions of spiritual or aesthetic value may resist being reduced down to a collection of variables, either because stakeholders are unwilling to do so, or are simply unable. Additionally, for some social objectives there are a wide range of perceptions and judgments of values making it difficult to boil down the model structure into something that is agreed upon by all stakeholders. The elicited probabilities may have so much uncertainty that is too great for a BN to give useful predictions. In order to answer these questions, further research is required on using BNs to predict social objectives.

ACKNOWLEDGEMENTS

This work was jointly funded by the UK Engineering and Physical Sciences Research Council and Economic and Social Research Council. We would like to thank the many canoeists and researchers in the URSULA project who have given time and help making this work possible.

REFERENCES

Ames, D.P., B.T. Nielson, D.K. Stevens and U. Lall, Using Bayesian networks to model watershed management decisions: an East Canyon Creek case study, *Journal of Hydroinformatics*, *7*(4), 267-282, 2005.

Barton, D. N., T. Saloranta, S. J. Moe, H. O. Eggestad and S. Kuikka, Bayesian belief networks as a meta-modelling tool in integrated river basin management -Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin, *Ecological Economics*, *66*(1), 91-104, 2008.

Borowski, I. and M. Hare, Exploring the gap between water managers and researchers: difficulties of model-based tools to support practical water management. *Water Resources Management*. 21(7), 1049-1074, 2007.

Gill, L., V. Kumar, E. Lange, D.N. Lerner, E. Morgan, D. Romano and E.A. Shaw, An interactive visual decision support tool for sustainable urban river corridor management. iEMSs, Ottawa, 2010.

Henriksen, H.J., P. Rasmussen, G. Brandt, D. von Bulow and F.V. Jensen, Public participation modelling using Bayesian networks in management of groundwater contamination, *Environmental Modelling and Software*, *22*(8), 1101-1113, 2007.

Holzkaemper, A., V. Kumar, B. Surridge, A. Paetzold, S. Lautenbach, R. Harris and D.N. Lerner, How to develop a simple tool for a complex task – Decision support for integrated catchment management, (submitted), 2010.

Kumar, V., A. Holzkaemper, B. Surridge, P.I. Rockett, M. Niranjan and D.N. Lerner, Bayesian challenges in integrated catchment modelling, paper presented at International Congress on Environmental Modelling and Software, Integrating Sciences and Information Technology for Environmental Assessment and Decision Making, Barcelona, Spain, July, 2008.

Kumar, V., A. Holzkaemper and D.N. Lerner, Capturing critical state effect in parameterisation of Bayesian network (in prep)

Millennium Ecosystem Assessment, *Ecosystems and human well-being: Biodiversity synthesis*, World Resources Institute, Washington, DC, 2005.

Soncini-Sessa, R., *Integrated and participatory water resources management; practice*, Elsevier, Amsterdam, 2007.