Using Bayesian Belief Networks to make integrated irrigation management decisions - A case study from Sri Lanka.

J.D. Cain[1], *R. Sakhtivadivel*[2], *I. Makin*[2], *C.H. Batchelor*[1]
[1]Institute of Hydrology, Wallingford, Oxon., OX10 8BB, UK.
[2]International Water Management Institute, Colombo, Sri Lanka.

Abstract

It is now widely accepted that successful irrigation management requires consideration of a number of physical, socio-economic and institutional factors. However, as these factors often interact in complex ways it is not always easy to do this. This paper presents an approach which is able to represent complexity in a simple, yet rigorous, way and facilitates integrated analysis of data from a wide range of sources. Bayesian Belief Networks are a statistical tool that have been widely used in disciplines such as medical diagnosis but have only recently been applied to environmental science and natural resource management. The general approach is presented, highlighting, in particular, how Belief Networks can be adapted to identify optimal decisions to management problems, and the advantages of taking this approach will be discussed. To illustrate their application, a case study from Sri Lanka is presented where a Belief Network has been constructed to help identify the most appropriate rehabilitation strategy for a tank cascade system. This network models the relative impacts of variable climate and human management skills on crop yield. Currently, the network represents each tank individually. Once fully developed, however, it will be able to recommend the optimal management strategy for the cascade as an integrated physical and socio-economic unit rather than as separate tanks. In conclusion, promising paths for future development of this approach will be outlined.

Introduction

Physical considerations of soil and water alone are not sufficient for effective irrigation management (Carter, 1993). Successful farmer irrigation is dependent on more than technical feasibility and will vary due to factors such as cost, a farmer's skill and the presence (or lack) of any communal arrangements. If there is sufficient water to meet the total demand, it is important that the effects of these, and other, socio-economic and institutional factors can be quantified. If there is not, then it is important to understand how they interact with the physical environment so demand can be managed effectively.

To facilitate this, a modelling framework is required which is capable of representing the inherent complexity which arises when the natural environment is considered in an holistic way. Moreover, the framework must enable the integrated analysis of different types of data measured at different scales and with (often large) associated uncertainties. Belief Networks (BNs) can provide such a framework. Based on Bayesian statistics, they were originally developed as a tool to aid decision analysis (Varis, 1997a, Varis and Kuikka, 1997). Subsequently, they have been applied to a number of different problems (Jensen, 1996), although only recently to the field of environmental modelling (Varis, 1997b, Stassopoulou *et al.*, 1998).

Their use provides a number of benefits :

- They allow simple conceptual representations of the environment to be converted into models which naturally integrate physical and socio-economic dynamics;
- The interaction between these dynamics can be defined with great flexibility and can be updated dynamically as the environment changes;

- The uncertainty in that interaction, and in the data used to specify it, can be explicitly incorporated into the model;
- Consequently, they provide an interface for the integrated analysis of both qualitative and quantitative data.

Moreover, as the approach is graphically based, people without specific skills can become involved in model construction and outputs are easily understood and communicated. This facilitates stakeholder involvement during the planning of management strategies and can help with subsequent implementation. In many ways, it is similar to other graphical modelling approaches (such as the Stella package), however, its statistical basis offers the ability to quantify uncertainty clearly and supports both predictive and diagnostic inference.

A case study is presented here which adapts a hydrological model developed by Sakthivadivel *et al.* (1997) to help with assessment of rehabilitation strategies for small tank cascades in Sri Lanka. In the original study, hydrological indicators estimated by the model were used in conjunction with non-hydrological criteria to select tanks suitable for rehabilitation. In the present study, by contrast, socio-economic and institutional factors are incorporated directly into the model within the framework provided by a BN, and different rehabilitation options are investigated in relation to crop yield for the cascade as a whole.

Constructing and using a Belief Network to make decisions

One of the strengths of using Belief Networks lies in their relative simplicity and the underlying functionality can remain opaque to the user. It is not the intention of this paper to explain the underlying theory but to demonstrate how the resulting tool can be applied to the problem of interest. A good introduction to the theory can, however, be found in Neapolitan (1990) and details of the statistical techniques are given by Spiegelhalter *et al.* (1993). The computational techniques used to calculate the statistical properties are outlined by Jensen (1996).

A Belief Network can be constructed in three stages. Firstly, the key variables which characterise the system behaviour are identified and assigned a finite set of states. Secondly, these variables are linked together to indicate cause and effect interactions between them. Thirdly, the function of each link is defined in a probabilistic manner. For example, we may wish to predict what the weather will like in the afternoon given the morning's weather forecast. The key variables in this system are identified as "Weather" (indicating the weather in the afternoon) and "Forecast" (indicating what the morning forecast predict the weather in the afternoon will be). The forecast is usually made in terms of whether it will be Sunny, Cloudy or Rainy so these are the states assigned to the Forecast variable. However, we are only really interested in whether it is going to rain or not in the afternoon so the Weather variable is assigned the states "No Rain" and "Rain". We assume that the forecast is based on some empirical approach so we draw an arrow linking Weather to Forecast to indicate that the weather, in some way, determines what the forecast is. This is shown in Figure 1 where the numbers to the right of each variable state indicate the probability that the variable is in that state (based on the information entered - see below). Finally, we define the functionality of the link by specifying probability distributions over the states of Forecast for each state of Weather. Table 1 indicates that, from past experience, when there is no rain in the afternoon, 70 % of the time the forecast was sunny, 20 % of the time it was cloudy and 10 % of the time it was rainy. A distribution for rain is specified in a similar way.

In this example, both variables have been given qualitative states as this most appropriate given the problem being considered. If it is required, however, variable states can be defined quantitatively. Moreover, the Network could be expanded to model the situation in a more

complex manner. For example, variables which influence the weather could be included although it might be assumed that these have already been measured and are inherent in the variable Forecast.

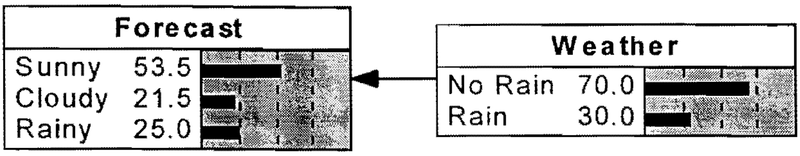


Figure 1 : Simple Belief Network representing a weather forecast model. The BNs shown throughout this paper are generated by the Netica software package produced by the Norsys Corporation.

Weather	Forecast is Sunny	Forecast is Cloudy	Forecast is Rainy
No rain	70	20	10
Rain	15	25	60

Table 1 : Conditional probability table specifying the link between Weather and Forecast in Figure 1

The BN can be adapted further so that it is able to make decisions concerning the system being modelled. When this is done it is more properly called a Decision Network (DN) or Influence Diagram. To extend the example above, we may wish to decide whether to take an umbrella or not on the basis of the weather forecast. We would, therefore include a decision variable in the Network called "Decide Umbrella" with states of "Take it" and "Leave it", and draw an arrow to it from Forecast.

The DN operates by weighing up the relative benefits of each possible outcome together with the chance that this outcome will be achieved. The relative benefits are measured in terms of dimensionless utility score which is quantified in the Network by a utility variable called "Satisfaction". Satisfaction is not just dependent on the decision taken but also on how the weather turns out. For example, if the umbrella is taken and it doesn't rain, satisfaction will be less than if the umbrella is left at home (because there is no need to carry the umbrella around all day). There is, therefore, also a link from the variable Weather to Satisfaction (Figure 2). The utility values for each possible outcome are shown in Table 2.

Based on the conditional probabilities shown in Table 1 and the utility scores shown in Table 2 the DN calculates the expected utility for each possible decision (Take it and Leave it). In Figure 2 the decision function for when the forecast is cloudy is shown. The values to the right of the states in the decision variable (Decide Umbrella) are the expected utility values. They are calculated by calculating the product of the utility value of each possible outcome with the probability of it being achieved and adding the relevant outcomes together. For example, the expected utility for Take it is calculated as (20 x 0.651) + (70 x 0.349) = 37.442. It is then assumed that the decision with the maximum expected utility is taken. In this case, it appears that leaving the umbrella is preferable to taking it.

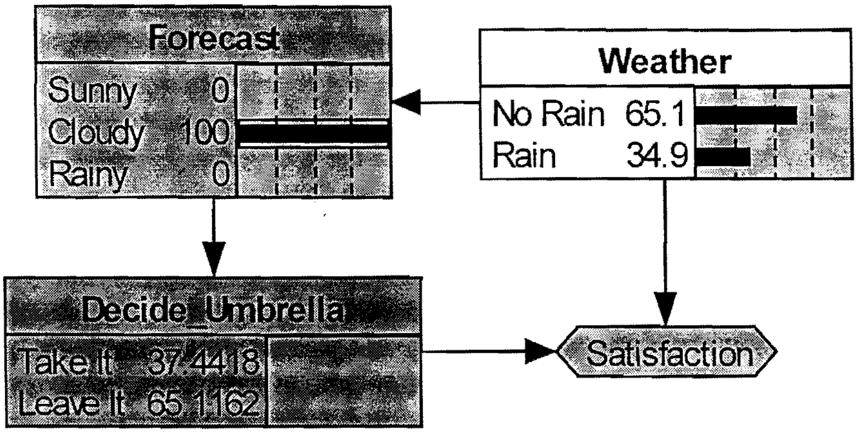


Figure 2 : Decision Network developed from Figure 1

Weather	Decide Umbrella	Satisfaction
No rain	Take it	20
No rain	Leave at home	100
Rain	Take it	70
Rain	Leave at home	0

Table 2 : Utility scores for the Satisfaction variable

The decision itself is made according to a rigorous statistical method and, consequently, there is no subjectivity involved. It is purely based on the values entered into the conditional probability and utility tables and once these are considered to be appropriate, the Network only serves to indicate the most logical choice. It must be stressed that this should not be taken as "the answer" to the management decision but only as a guide as to what the best choice might be.

While the description above may sound involved, it should be noted that once the user is familiar with the software and the underlying theory, constructing a DN and the interpreting its outputs becomes a relatively rapid procedure. For example, the whole Network development process, from design through data collection to construction, took a little over three weeks.

Development of the tank cascade model

In their original study, Sakthivadivel *et al.* (1997) based decisions regarding tank rehabilitation on the hydrology of the cascade, the number of beneficiaries, the rehabilitation history and the financial investment. For this study, their approach was adapted in two ways :

1. The hydrological model was set into a BN framework and additional variables representing socio-economic factors were included;

2. The BN was extended (to form a DN) to make decisions based on the total crop yield produced and the financial investment.

The BNs for each of 12 tanks in the cascade were constructed separately. Variables were assigned to each element of water balance and given quantitative states appropriate to the values they were expected to take. The purely hydrological relationships (e.g. between runoff and the net water in the tank) were specified according to values used in the original study, although the probabilistic nature of the BN enabled any uncertainty in these data to be explicitly incorporated.

The additional socio-economic factors represented the management skill of the farmers using water from each tank and the effectiveness of local institutions in ensuring equitable access to the water. These were added as variables in the Network. Further biophysical factors were also added, namely the state of the water distribution infrastructure and the effect of pest damage. These factors were chosen both because they were considered to have a strong effect on the yield and because data describing them were available. All these variables were assigned qualitative states reflecting the data available. Given the inexact nature of some of these variables there was uncertainty attached not only to the values assigned to them by measurement but also to how their interactions could be modelled (for example, the relation between overall water management and crop yield is difficult to characterise). To address this, estimates of the uncertainty arising from these considerations were included in the conditional probabilities derived from the measured data. The completed Network is shown in Figure 3.

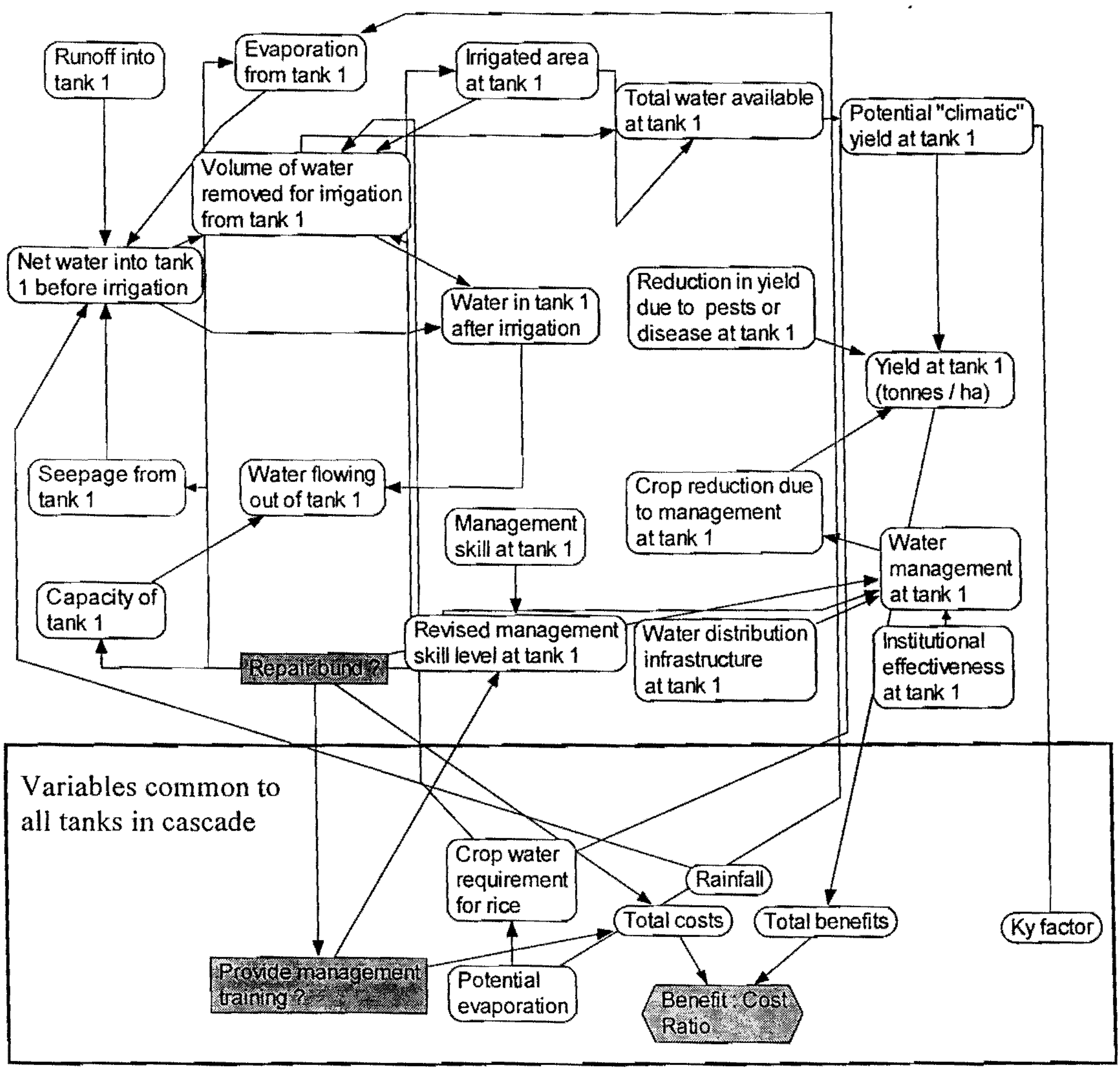


Figure 3 : Decision Network for tank 1

Decisions were added to the Network constructed for each tank on the basis of those identified to be most appropriate by the farmers themselves in the case of Tank 1, the option of repairing the bund was highlighted by the farmers). In addition, a decision regarding the provision of management training was included for each tank to evaluate whether extension advice would significantly contribute to increased crop yields.

It is intended that the DNs developed for each tank be linked by, for example, joining the variables "Water flowing out of tank 1" and "Water flowing out of tank 2" to the variable "Net water into tank 3 before irrigation" (as this is the cascade structure). In addition, the variables at the bottom of Figure 3 will be linked to all tank Networks. Currently, difficulties are being encountered with computing power due to the resultant size of a Network which consists of 12 repeated modules. Experience with other DN applications suggests that this level of complexity is rarely necessary and various optimisation and approximation techniques are currently being investigated to provide a solution when it is. Once a suitable solution has been identified, however, the Network will be able to make decisions which optimise the benefits for the catchment as a whole rather than for each tank individually.

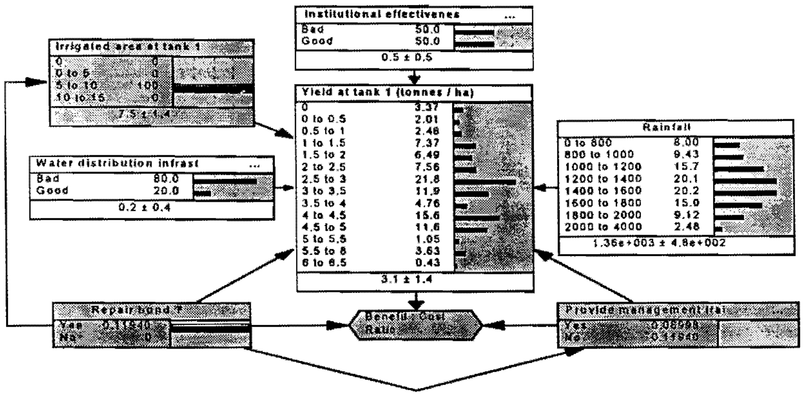
Investigating different rehabilitation strategies

Once compiled, the Network immediately provides the expected utilities for each possible decision for each decision node. The results for tank 1 are shown in Figure 4 for the case where the irrigated area of tank 1 lies between 5 and 10 hectares and all other variables are uncertain (the figure shows only a portion of the Network). By identifying the maximum expected utility value, it can be seen that tank 1 yield is optimised by choosing to repair the bund and not providing management training. The result from the "Repair bund" decision variable is obvious as the yield is modelled as that produced from irrigation. Consequently it is zero of the bund is breached and no irrigation water is available. The result from the "Provide management training" variable is less obvious, however, it indicates that the benefits to be gained by providing training are less than the costs of doing so. This would be expected to change when the whole cascade is considered as the cost of training provision would not increase greatly above that for one tank while the benefits gained would be multiplied by the number of tanks in the cascade.

It can seen from Figure 4 that the uncertainty in the yield is high, as indicated by the relatively flat probability distribution across the states and the large error given in the expected yield of 3.1 tonnes / ha. This is due to the combination of uncertainties associated with all the Network variables. While this may not provide a conclusive indication of the yield, it is useful in that it provides the decision maker with a real evaluation of the risk associated with a particular decision arising from imperfect knowledge. As a decision maker will never have perfect knowledge this is an important feature of the Decision Network. Moreover, further analysis of the Network can reveal the most useful variables for which to collect new data in order to reduce the uncertainty in the yield. Any new data acquired can be used to "teach" the Network and improve its functional relationships, without wholly rejecting the data previously entered.

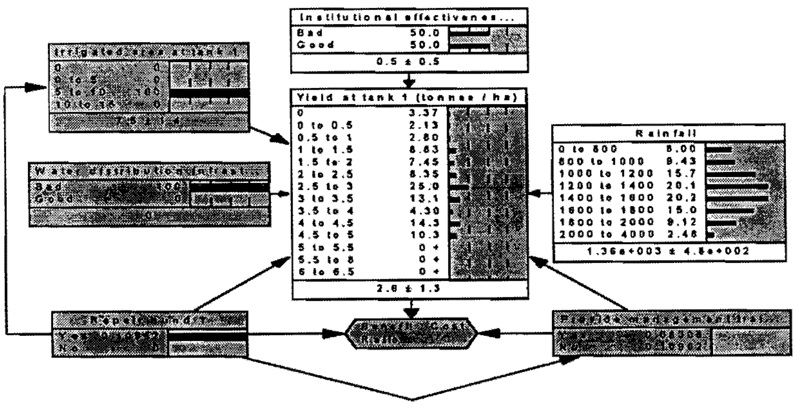
The Network can also be used to examine different "what if ?" scenarios. For example, it may be of interest to know how repairing the water distribution infrastructure (a option not originally proposed by the farmers) may effect the yield and the optimal decisions. Figure 5 shows the effect of assuming that the infrastructure is "Good" (by selecting that state in the Infrastructure variable) as compared to assuming that it is "Bad". It can be seen that the choice of optimal decision is unaffected, although the expected yield increases from 2.8

tonnes / ha to 4 tonnes / ha between the different situations. All the other variables in the Network can be altered in a similar way, either singly or in combination, to investigate a



whole range of possibilities.

a



b

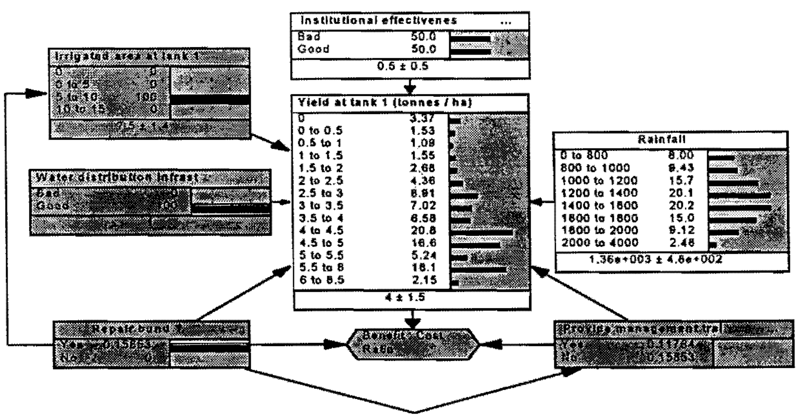


Figure 4 : Compiled Decision Network for Tank 1 (only a portion of the DN is shown)

Figure 5 : Compiled Decision Network for Tank 1 showing the relative effects of (a) bad and (b) good water distribution on yield

Conclusions

The case study presented here suggests that Belief Networks provide an effective means developing integrated models to support irrigation water management. They facilitate the understanding of complex environmental interactions by providing a simple, yet logical, framework within which to work. Moreover, as a common interface for multi-disciplinary data analysis, they enable socio-economic and institutional considerations to be considered on an equal basis with relevant physical factors.

They can be adapted to identify optimal management decisions based on explicit incorporation of the uncertainties associated both with data error and the relationship between variables. They hence give the decision maker a realistic assessment of the risks associated with a particular decision for a variety of scenarios. This information provides a valuable foundation on which the best decision can be made.

The approach is still relatively new to the environmental sciences and much further work is required. In particular, there is a need to identify rigorous methods for quantifying and combining the uncertainty which arises from different sources. This necessarily involves a consideration of the most appropriate scales to work at and how data measured at different scales can be combined. As decisions are taken with reference to some goal, appropriate indicators must be found which define whether that goal has been achieved. Indeed, if DNs are to be used most effectively, thought must be given not only to what data are collected but how these data are elicited. This is particularly important where data seeks to quantify social dynamics.

Acknowledgments

The funding for this work was........

References

Carter, R. 1993. Small-scale irrigation in sub-saharan Africa : a balanced view. Proceedings of the Conference on Priorities for Water Resources Allocation and Management, Natural Resources and Engineering Advisers Conference, Southampton, July 1992. Overseas Development Administration, London, 103-116.

Jensen, F.V. 1996. An introduction to Bayesian Networks. 1st Edition, UCL Press, London, 178 pp.

Neapolitan, R.E. 1990. Probabilistic reasoning in expert systems : Theory and algorithms. John Wiley and Sons, New York.

Sakthivadivel, R., Fernando, N., and Brewer, J.D. 1997. Rehabilitation planning for small tanks in cascades : A methodology based on rapid assessment. Research Report 13. Colombo, Sri Lanka, International Irrigation Management Institute, 13 Edition, 32 pp.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L., and Cowell, R.G. 1993. Bayesian analysis in expert systems. Statistical Science, 8 (3) : 219-283.

Stassopoulou, A., Petrou, M., and Kittler, J. 1998. Application of a Bayesian network in a GIS based decision making system. International Journal of Geographical Information Science, 12 (1) : 23-45.

Varis, O. 1997a. Bayesian decision analysis for environmental and resource management. Environmental modelling and software, 12 (2-3) : 177-185.

Varis, O. 1997b. Bayesian decision and risk analysis in environmental assessment. In : Zannetti, P. (Ed.). Environmental modelling, Vol. IV. CMP/Elsevier, Southampton,

Varis, O. and Kuikka, S. 1997. Joint use of multiple environmental assessment models by a Bayesian meta-model : the Baltic salmon case. Ecological Modelling, 102 : 341-351.