Management of the Environment and Resources using Integrated Techniques

Guidelines for the use of Bayesian networks as a participatory tool for Water Resource Management

Compiled by John Bromley

July 2005

Centre for Ecology and Hydrology
Maclean Building
Crowmarsh Gifford
Wallingford
0X10 8BB

Table of Contents

Appendices

Appendix I Basic statistical background of CPTs and properties of Bns

Appendix II Description of how to complete conditional probability tables: Large data sets and model output.

Appendix III Descriptions of how to complete conditional probability tables: Stakeholder or expert opinion

Compact Disc

Folders:

UK: UK networks plus documents

Denmark: Danish network plus documents

Spain: Spain networks plus documents

Italy: Italian networks plus documents

Hugin Viewer v 6.5: Hugin Program

Merit guidelines: Copy of the Merit guidelines

CPT calculator: J.Cain's calculator using EPTs to complete CPTs

JCain_Guidelines: Copy of the Cain (2001) guidelines

About the guidelines

These guidelines describe the way in which Bayesian networks (Bns) can support decision making in the field of water resource management. Though primarily intended for the water resource sector, they could just as easily be applied more widely to the sphere of environmental management. The guidelines have two objectives:

- The first is to describe the way in which Bns can be used to assess the impact of any water resource management decision, not just in terms of the water resource itself, but also the effect it has on the social and economic life of the region. By linking environmental, social, economic and other types of factor in this way, the technique permits a genuinely 'integrated' evaluation of any particular decision.

- The second objective is to show the way in which Bns can be used to help involve stakeholders more fully into the decision making process by providing a focus for the stakeholder engagement process. Issues and objectives identified by stakeholders can be incorporated into networks constructed with their full involvement. Moreover, the graphical nature of network presentation makes it easier to illustrate the impacts of a range of different management strategies to stakeholders.

However, it should be stressed that in these guidelines networks are not used to make decisions, but simply to provide the information on which better-informed decisions can be made.

The guidelines are presented in 6 sections with Appendices and a CD (enclosed in a pocket in the back cover). The CD includes working copies of case studies, a PDF copy of the guidelines, along with other papers and reports produced during the project. There is also a copy of the report by Cain (2000). For a full list see the Table of Contents.

Preface

The Management of the Environment and Resources using Integrated Techniques (MERIT) is a research project supported by the Fifth European Community Framework Programme for Research and Technical Development, under the call 'Energy, Environment and Sustainable Development (1998-2002)'.

Contract Number EVK1-CT-2000-00085 MERIT was signed by DGXII in May 2001 and began on 1st June 2001. This report synthesises the work undertaken from 1st June 2001 to 31st May 2004. The MERIT team comprised:

1. **Centre for Ecology and Hydrology - (CEH):** *John Bromley, Mike Acreman, Nick Jackson and Anna Maria Giacomello,*
2. **Aalborg University (AU):** *Finn Jensen*
3. **Geological Survey of Denmark and Greenland (GEUS):** *Per Rasmusen, Hans Jørgen Henriksen, Gyrite Brandt*
4. **Universidad de Castilla la Mancha (UCLM):** *Francisco Martín de Santa Olalla Mañas, Alfonso Domínguez Padilla, Fernando Ortega Alvarez, Alfonso Artigao, Concha Fabiero*

5 Sperimentale Italiano Giacinto Motta S.p.a (CESI): *Guiseppe Paolo Stigliano, Stefano Maran*

6 Politecnico di Milano (DEI): *Rodolfo Soncini-Sessa, Andrea Castelletti, Enrico Weber*

7 Regione Abruzzo Giunta Regionale: *Angelo D'Eramo, Natalia O. Pisegna*

8 University of Birmingham Centre for Environmental Research & Training: *Judith Petts, Olivia Clymer, John Gerrard*

The project was undertaken with the active co-operation and collaboration of four organisations responsible for water resource management in the selected study sites. The four organisations were:

i. The Environment Agency (UK)
ii. Copenhagen Energy (Denmark)
iii. Regione Abruzzo Giunta Regionale (Italy) – who also form part of the funded team
iv. Junta Central de Regantes de La Mancha Oriental (Spain)

The programme co-ordinator was John Bromley; meetings were chaired by Mike Acreman.

Acknowledgements

This work was carried out as part of the European Union Fifth Framework programme, specifically under the key action "Sustainable Management and Quality of Water" which lies within the Energy, Environment and Sustainable Development Thematic Programme. The MERIT partners would like to thank Negjmija Rugova for overseeing the financial and administrative aspects of the project and Dr. Panos Balabanis for his encouragement and support in his role as the responsible EU scientist. Special thanks are also due to all the organizations and agencies that co-operated so enthusiastically with the MERIT team to make the project a success. These organizations include the Environment Agency (UK), Copenhagen Energy (DK), the Junta Central de Regantes de la Mancha Oriental (ES), and the Giunta Regionale L'Aquila (IT).

Many thanks also to our two external reviewers; Dr. Patrick Moriarty from the International Water and Sanitation Centre (IRC), Delft, Netherlands, and Dr John Butterworth from the Natural Resources Institute (NRI), Chatham, UK. They both provided constructive comments on the science and project management.

Finally it should be stressed that much of the inspiration for these guidelines stems from the work of Jeremy Cain, who pioneered the combined use of Bayesian networks and stakeholder participation techniques for water resource management as part of a project funded by DFID [1] (Cain, 2001)[2].

[1] Department For International Development (UK).

[2] Planning improvements in natural resources management; August 2001. Published by CEH-Wallingford. This report is available on the MERIT website www.merit-eu.net, and on the CD accompanying these guidelines.

1. INTRODUCTION

1.1 Background and relevance to the Water Framework Directive

In recent years the provision of a reliable and safe water supply for people worldwide has become a major issue on the international agenda. To highlight concern the year 2003 was declared by the UN to be the 'International Year of Freshwater'. At the launch it was stated that:

"…….*By developing principles and methods to manage this resource* (water) *efficiently and ethically, while respecting related ecosystems, we move a step closer to the goal of sustainable development."*

This statement reflects acceptance by the international community that the long term viability of water supplies throughout the world can only be secured through *sustainable* development of the resource. Moreover it is generally accepted that the best way to achieve sustainable development is by implementation of an Integrated Water Resource Management (IWRM) approach. The concept of IWRM is based on principles of good water management drawn up over a number of international meetings, the most important of which include:

- United Nations Water Conference in Mar del Plata (UNDP, 1977)
- International Conference on Water and Environment in Dublin (WMO, 1992)
- The Earth Summit in Rio de Janeiro (UNDP, 1992)

These are often referred to as the Dublin Principles (Box 1.1).

Box 1.1 The Dublin Principles

- *Freshwater is a finite and vulnerable resource, essential to sustain life, development and the environment;*
- *Water development and management should be based on a participatory approach, involving users, planners and policy-makers at all levels;*
- *Women play a central part in the provision, management, and safeguarding of water;*
- *Water has an economic value in all its competing uses, and should be recognised as an economic good.*

In 2000 the European Commission endorsed the principles of sustainability and IWRM through enactment of the Water Framework Directive (WFD) (Directive 2000/60/EC, October 2000). This obliges all Member States to produce River Basin Management Plans by 2010. These are to be implemented using the principles of IWRM, which include the requirement for public participation in the process. Although the phrase 'public participation' does not appear in the Directive itself, the EU has produced a guidance document on this subject under the Common Implementation Strategy[3]. In this document three increasing levels of public involvement are considered:

[3] Public Participation Techniques, August 2002. Download on: http://www.wrrl-info.de/docs/Annex1.pdf

- Information supply
- Consultation
- Active involvement

According to the guidance the first two are obligatory, the latter encouraged. The implications of the WFD are that for the first time water management in Europe will need to demonstrably integrate analysis with the genuine participation of the wider community in the decision making process.

Although there is now international acceptance of IWRM as the best way to achieve sustainable development the application of these principles is still far from being widely applied. This is partly because more time is needed to change from a long-established sectoral approach, but also because the tools for implementation have not yet been fully developed; much needs to done before IWRM becomes a reality.

By requiring the implementation of IWRM the Water Framework Directive has thrown down a challenge to the Community: namely to find the tools, techniques and methodologies that will enable successful implementation of integrated management and thereby secure a sustainable future for the continent's water resources.

The present guidelines describe one technique that may help facilitate the integration process and increase the degree of stakeholder participation required for success. The technique is a decision support tool that makes it possible to evaluate the interaction and links between different types of factor in a genuinely holistic way, and to provide an output on which it is possible to base a fair and balanced decision. Moreover the method encourages the active involvement of stakeholders, whose opinions and input are sought throughout the entire process.

The methodology is based on the use of Bayesian networks to link together the various factors that need to be considered when a management decision is being made. Through a process of carefully organised consultation, stakeholders are able to influence the design and content of the networks and in some cases even provide data.

1.2 Integrated Water Resource Management (IWRM): what is it?

There are numerous definitions of IWRM (Box 1.2), and although each may have a different emphasis, two common threads run through them. The first is that management must be interdisciplinary. This means that it is not sufficient to simply evaluate the impact of a particular strategy on the water resource alone; it must also take into account any social, economic, environmental or cultural effects the policy may have. The second is that decisions must involve the participation of members of the community affected by the strategy, in other words the stakeholders – that is to say the most affected members of the community, rather than just the most powerful and organised, or only the legally involved parties.

But what does integration really entail? It means that the impact of management decisions must not be restricted to the water resource itself, but also embrace the wide range of other factors that play a role in the life of a river basin. For instance, the decision to build a dam, which may form part of a River Basin Management Plan, will have an immediate impact on the water resource of the region. But there will also be wider social, economic, and perhaps political repercussions. Some sections of the population may benefit, others may suffer.

To make a balanced and fair judgement we need to be able to evaluate the impacts of decisions on an extensive array of factors. Many of these impacts will conflict. A dam will provide more water, but result in the loss of land and housing; it may provide recreational facilities, but deny water to communities further downstream. Dams are also expensive to construct and maintain, so the economic implications of construction need to be considered as well. An integrated policy requires all these benefits, drawbacks and costs to be considered and evaluated. Inevitably the issues involved will be many and varied; some will be physical (e.g. river flow; groundwater levels), others economic (e.g. agricultural output), or social (e.g. recreational facilities) or of some other type.

Box 1.2 Definitions of Integrated Water Resource Management (IWRM)

- *IWRM is a process which promotes the co-ordinated development and management of water, land and related resources, in order to maximise the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems (Global Water Partnership, Technical Advisory Committee, 2000)*[a]*.*
- *Integrated Water Resource Management (IWRM) is simultaneously a philosophy, a process, and an implementation strategy to achieve equitable access to, and sustainable use of, water resources by all stakeholders at catchment, regional, national, and international levels, while maintaining the characteristics and integrity of water resources at the catchment scale within agreed limits (DWAF 2003)*[b]

[a] www.gwpforum.org
[b] www.dwaf.gov.za

Of course a dam may not be the only option available to the planner; alternatives may include the installation of a well field, the import of water from outside the region, water demand control measures, or indeed a combination of all these. In this case the impact of the different options must be evaluated before a decision can eventually be made. The problem is to find a way of evaluating everything that needs to be considered for each of the different options, and eventually reach a decision that is *equitable, efficient* and *sustainable*. This is the type of complex problem the implementation of IWRM poses, but one which must be addressed if the policy is to be successful.

There is also the question of the nature of the stakeholder involvement in the decision making process. In the context of IWRM, involvement does not simply mean consulting the community once decisions have already been made, but rather actively involving a representative cross section of organisations and individuals in the decision making process itself. It is important for consultation to start at the very outset, so that stakeholders are given the opportunity to identify the issues that are important to them, and to have a say in what options might be considered. This gives the opportunity for people with different points of view to express their opinion and to introduce local knowledge into the discussions. Conflict may result, but part of the integrated approach is to incorporate techniques of achieving consensus. Failure to secure full involvement and support of stakeholders will ultimately prejudice successful implementation. The problem is to find techniques to involve the community that are transparent, as well as practical and effective.

1.3 Approaches to integration: tools, techniques and models

As we have seen the task of evaluating different management options within an environmental-social-economic system to obtain a decision that is equitable, efficient and sustainable, is complex. To help them, planners and decision makers need a range of tools, techniques and models. But because there is no simple blueprint for the application of IWRM, each case will be different, and the type of tool or model required may also be different. It is up to the planner to select the best approach. To aid the implementation of IWRM a portfolio of policies and approaches has been assembled by the Global Water Partnership[4] (GWP) into a 'ToolBox' that is available on a Web site (www.gwpforum.org) (Box1.3). The toolbox is organised under 3 main headings:

- The Enabling Environment - legal, policy and financial aspects
- The Institutional Roles - the structure of controlling organisations
- Management Instruments - techniques to control water supply and demand

Box 1.3 Global Water Partnership IWRM ToolBox

The main headings in the ToolBox

A The Enabling Environment

A1 Policies
A2 Legislation
A3 Financing and Incentive Structures

B Institutional Roles

B1 Creating an Organizational Framework
B2 Institutional Capacity Building

C Management Instruments

C1 Water Resources Assessment; (C.1.3. Models and Decision Support Systems)
C2 Plans for IWRM
C3 Demand Management
C4 Social Change Instruments
C5 Conflict Resolution
C6 Regulatory Instruments
C7 Economic Instruments
C8 Information Management Exchange

(www.gwpforum.org)

The first two headings represent legal, financial and institutional actions and reforms that by and large need to be done at the regional and national levels to provide the overarching framework within which IWRM can be successfully implemented. The third heading, however, deals with the management instruments available to tackle the problem at the catchment scale and below. The management technique being developed by MERIT falls under item C.1.3 entitled 'Modelling in IWRM'. Under this item models and decision support systems are identified as important management tools.

A model is a simplified description of a system to assist calculations and predictions. Many models have been designed to facilitate integration between various aspects of

[4] The GWP is a working partnership among those involved in water management: government agencies, public institutions, private companies, professional organizations, multilateral development agencies and others committed to the Dublin-Rio principles.

catchment hydrology, including surface water, groundwater, vegetation, ecology, and even agricultural economics. Examples include NELUP (O'Callaghan, 1995) MIKE SHE (Refsgaard and Storm, 1995), and TOPOG (Vertessy et al., 1994). A description of many others can be found in Singh (1995). These types of model are excellent for water resource assessments and impact on the environment, but in most cases they do not link directly to the wider social, cultural and economic aspects of water management.

Decision support systems (DSSs) are complementary to models. A DSS is a means of collecting data from many sources to inform a decision. Information can include experimental or survey data, output from models or, where data is scarce, expert knowledge. On their Web site the GWP state that a DSS should allow users to integrate data in 5 phases, each requiring consultation with stakeholders:

- *Issue identification*
- *Defining management options*
- *Establishment of decision criteria*
- *Data acquisition*
- *Decision support process*

There are many types of decision support system, good descriptions of some of the more up to date techniques being given in Marakas (1999), and Turban and Aronson (2000). Cain (2001) identifies a number of the more widely used types of DSS and lists some of the associated commercial packages; the types include influence diagrams, decision trees, mathematical models, multi-criteria analysis and spreadsheets.

These guidelines describe the application of another type of decision support tool to the field of water resource management, *Bayesian networks* (Bns). For many years Bns have been used routinely in the fields of medicine and artificial intelligence, but until now have had limited application to environmental problems. We have used a Danish commercial software package produced by HUGIN (*www.hugin.dk),* although other software is also available, such as that produced by Netica (*www.norsys.com*), BayesiaLab (www.bayesia.com) and Bayesware Discoverer (www.bayesware.com). A comprehensive list of available Bn software, together with details of their functionality, can be found at (*http://www.ai.mit.edu/~murphyk/Software/BNT/bnsoft.html).*

1.4 The MERIT project

At this point a brief description of MERIT[5], the project on which these guidelines are based, is presented. The 3 year project, funded by the EU, and completed in 2004, aimed to develop an integrated, water resource management methodology that could be applied throughout Europe.

The methodology that has been developed exploits the flexibility of *Bayesian networks* (Bns), a type of decision support system used successfully for many years in fields such as medicine and artificial intelligence. Bayesian networks simulate the operation of natural systems, and are most effective when designed and set up with help from the people who live and work in that system, the ‘stakeholders’. The

[5] MERIT: The **M**anagement of the **E**nvironment and **R**esources using **I**ntegrated **T**echniques is a research project supported by the Fifth European Community Framework Programme for Research and Technical Development, under the call ‘Energy, Environment and Sustainable Development (1998-2002)’. Contract Number EVK1-CT-2000-00085

methodology, therefore, places great stress on the need for local communities to contribute toward the construction of these decision support systems. When all's said and done, it is their lives that are affected by subsequent management plans and decisions.

MERIT had two main objectives: (a) to demonstrate the extent to which networks can be an aid to water management decision making and (b) to investigate whether Bns can be useful as a tool to promote stakeholder involvement. Networks were set up in full consultation with local stakeholder groups and organisations**.** To develop the methodology, case studies were undertaken in four catchments throughout Europe, in the UK, Denmark, Spain and Italy, representing a range of environments, subject to different pressures and in contrasting social settings.

The issues varied from country to country:

- <u>UK, Loddon Catchment</u>

Water Demand Network: This network focused on the management of domestic water demand in the Loddon, a catchment in south east England. It was carried out in collaboration with the Environment Agency, the main regulatory water authority for England and Wales.

- <u>Denmark, Havelse Catchment</u>

In Denmark two networks were developed:

(a) *Compensation Network:* A network constructed to investigate the potential of compensation payments to encourage farmers to reduce or stop the use of pesticides in the Havelse catchment in North Zeeland. Excessive application of pesticides has led to the pollution of the groundwater supply to Copenhagen.

(b) *Flooding Network:* A second network looks at the potential flooding problem in the lower part of the Havelse, resulting from the creation of new wetlands in the area.

In both cases the networks were set up in collaboration with Copenhagen Energy, the private company responsible for the supply of water to Copenhagen.

- <u>Italy, Vomano Catchment</u>

Again, two networks have been developed:

(a) Hydropower-Irrigation Network

This network deals with the increasing water requirements for irrigation in the Vomano catchment in the face of competing demands for hydropower. A network, to simulate the behaviour of farmers in the lower part of the Vomano River in east central Italy, was developed to help estimate irrigation water requirements in the region. The network was then linked directly to an existing operational reservoir management model, designed to calculate water requirements for hydropower generation. The linked model was used to select an operational schedule acceptable to both the farmers and the hydropower company. This work was carried out with the full cooperation of Sperimentale Italiano Giacinto Motta S.p.a (CESI), the company responsible for hydropower generation in the region.

(b) Environmental Flows Network

A second network was developed by CESI to investigate the ‘environmental flows’ needed to satisfy different groups of stakeholders in the area. Environmental flow is the minimum flow of water needed to sustain specified aquatic species.

- Spain, Júcar Catchment

The Spanish network deals with the intense competition for water between domestic, environmental and agricultural sector requirements in the Júcar catchment, in the eastern central part of the country. Excessive consumption, particularly for irrigation, has led to a dramatic decline in groundwater levels and a resulting unsustainable situation in the East La Mancha aquifer. The network examines the likely impact of various management interventions on the different stakeholder groups in the region. The network has been constructed with the active support and encouragement of the Junta Central de Regantes de La Mancha Oriental, the organization responsible for water management in the Júcar.

The key point is that each of the above networks have been developed in full consultation with stakeholders and end users, the aim being to develop a decision making process that is fully inclusive and transparent.

1.5 Structure of the guidelines

The remainder of the document is structured as follows:

***Section 2: Bayesian networks**:* This section describes the main features of Bayesian networks, but does not delve deeply into the statistical background. The aim of the section is simply to provide sufficient detail to allow a non-expert to understand the principles of the technique and begin construction of their own networks.

Section 3: Principles of stakeholder engagement: Here, the principles and methods of stakeholder engagement are discussed in the context of Bn construction, and from the wider standpoint of river basin management. Specific recommendations of what should and shouldn’t be done are given, along with more general guidelines concerning the requirements of a successful stakeholder engagement process. Examples are drawn from the MERIT case studies.

Section 4: Building the model, a guide to network construction: A detailed seven-stage guide to model construction and stakeholder engagement is presented. Each stage is described under two headings: 1) Network construction, dealing with the technical issues related to the design and creation of the network itself, including data entry, and 2) Stakeholder involvement, which describes the way in which stakeholders can and should be engaged in the process of network construction along each step of the way.

Section 5: MERIT Case Studies: The section provides a brief description of each of the 4 case studies. Working copies of each network are included on the CD in the back pocket of the guidelines.

Section 6: Bayesian networks, strengths and weaknesses as a participatory water resource management tool: A ‘Strength, Weakness, Opportunity and Threat’ (SWOT) analysis of networks as a participatory tool is presented in tabular form.

Appendices

Appendix I: Basic statistical background of CPTs and properties of Bns

Appendix II: Description of how to complete conditional probability tables: Large data sets and model output.

Appendix III: Description of how to complete conditional probability tables: Stakeholder or expert opinion

CD: A CD is included in the pocket at the back of the guidelines. This contains copies of all the case study networks, a PDF version of the guidelines, plus a number of other publications and reports. For details see the Table of Contents and accompanying CD.

2. BAYESIAN NETWORKS

2.1 Introduction

Over the past few decades various types of decision support system have been developed and applied to the complex field of environmental management. Generically known as Environmental Decision Support Systems (EDSS), these techniques encompass a wide range of technologies, including mathematical modelling, optimisation, geographical information systems, Monte Carlo simulation and so on (Huber, 1997; Jain et al., 1998).

The difficulties of environmental management in general and water resource management in particular, have recently been compounded by the requirements of legislation such as the Water Framework Directive; management strategies now need to be not only effective but demonstrably integrated, sustainable and participatory. This means that the impact of any decision has to be evaluated, not simply in terms of the water resource, but also in terms of the ecological, economic, social, cultural, and other effects it may exert.

There are many types of EDSS generator: influence diagrams, decision trees, mathematical models, multi-criteria analysis, spreadsheets, Bayesian networks and others. But which of these is most appropriate for the problem in hand, or indeed if any method is required at all, needs to be given careful consideration at an early stage in the decision making process. Each approach has its strengths and weaknesses, and selecting one that is inappropriate can be costly in terms of both time and money. These guidelines describe the way in which Bayesian networks may be used as a participatory EDSS, and under what conditions they may or may not be appropriate.

2.2 What is a Bayesian Network?

The first thing to stress is that you don't need to be an expert in mathematics or statistics to use Bayesian networks. Commercial packages are available that enable networks to be constructed and data entered without knowledge of the underlying equations. Nevertheless it is obviously beneficial to understand something of the general principles on which networks are based, and to be aware of what is, and what is not, possible. The aims of this chapter are to explain in simple terms the basic rules for the use of Bns, and to show how they can be used as an aid to environmental decision makers.

A Bayesian network is a type of decision support system based on a theory of probability using Bayes'[6] rule; this rule describes mathematically how existing beliefs can be modified with the input of new evidence. As more data about an environmental system becomes available, our understanding of the system and how it will respond to future impacts will (or may) be changed. The practical application of Bns is not new. They have been used successfully for many years in the fields of medical diagnosis and artificial intelligence, but to date their use in environmental systems has been limited (Varis, 1995; Varis, 1998). Part of the reason for this is that it is only in recent years that PCs have become powerful enough to cope with the large amounts of data and calculations involved with complex systems.

Bayesian networks are based on the concept of conditional probability. This is illustrated by a simple example given in Figure 2.1. In this example we suppose that the state of the factor 'Annual River Flow' depends on the states of two other factors 'Annual Rainfall' and '%Forest Cover'. This means that if either the forest cover or the rainfall, or both, are changed then river flow will also change. River flow is said to be conditionally dependant on the states of forest cover and rainfall. In the diagram the link and direction of the link between the factors is indicated by arrows. These arrows show the direction of the cause and effect; thus while changing the forest cover and/or annual rainfall can affect annual river flow, the reverse is not possible.

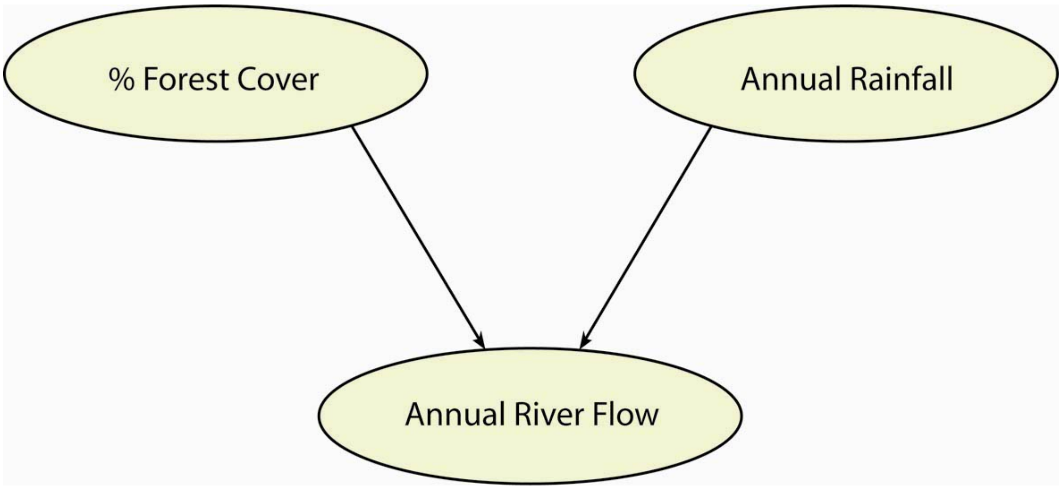


Figure 2.1 Conditional probability

A Bayesian network comprises three elements; firstly a set of variables that represent the factors relevant to a particular environmental system or problem, secondly the links between these variables, and finally the conditional probability tables (CPTs) behind each node that are used to calculate the state of the node. The first two elements form a Bn diagram; the addition of the third forms a fully functioning network. In this way Figure 2.1 is a Bn comprising 3 variables, represented as 'nodes', and two 'links' shown as arrows.

2.3 The components of a network

2.3.1 Nodes

[6] Thomas Bayes was an 18th Century English clergyman

Once identified, the variables considered to be relevant to a particular problem are represented in a network as nodes (the term 'nodes' and 'variables' are used interchangeably throughout these guidelines). As already mentioned these factors can be of any type; physical, environmental, social, and so on. Once identified each node is assigned a series of 'states', which represents the range of conditions that the node might potentially occupy under different conditions. These states can either be descriptive or numerical, depending upon the requirements of the decision maker and the nature of the node itself. There is no limit to the number of states that may be applied, though the more states used the more complex will be the resulting CPTs that describe the relationship between nodes (see section 2.3.3). With respect to the allocation of states there are two types of node; 'discrete chance' and 'continuous chance' nodes. For discrete nodes, the potential range of conditions is described as a series of separate states, whereas the states in continuous nodes are given as a Normal (Gaussian) distribution function, described in terms of its mean and variance.

Where discrete nodes are used, the states can be described in any one of four ways:

1) As a set of <u>labels</u>; e.g. high, medium, low.
2) As a set of <u>numbers</u>; e.g. 10, 20, 30
3) As a set of <u>intervals</u>; e.g. 0 - 5, 5 -10, 10 -15
4) In <u>Boolean</u> form; e.g. true, false

Examples of all these types of are given in section 4.6, which describes the procedure for assigning the states of variables.

In the case of continuous nodes, the state is described as a 'Gaussian (normal) conditional distribution function', in which the user defines the mean and variance rather than a set of specific states. There are, however, a number of restrictions that apply to the use of this type of node, and as a result are not widely used.

2.3.2 Links

The way in which nodes are linked within a network is decided by the user, though where the data permits the software itself can be used to suggest links. But by and large it is the user who decides on the pattern of links based on their experience and knowledge of the system and on information collected from other people (e.g. stakeholders). This is a crucial stage in network development and is discussed in more detail in section 4.4. The rules that govern the use of links are described in Appendix I.

The links represent causal relationships between nodes. The direction of the link is from cause to effect. Where two nodes are linked the destination node is termed the 'child', the node from which the link originates is known as the 'parent'. Where a node has no links from any other variable, the user is expected to define the state of the node. These parentless variables can be used to represent: (a) a possible action (b) a scenario that might arise or (c) an observed (known) condition. Where they are used to represent different scenarios we can call them 'scenario defining nodes' (Bromley et al., 2004).

To illustrate the concepts of nodes and links a basic network which has 9 nodes and 8 links is shown in Figure 2.2. This example represents the effects of changing the forest cover in a catchment where we are interested in the impact of these changes on farmers' income, the operation of a local reservoir, and the angling interests along the river.

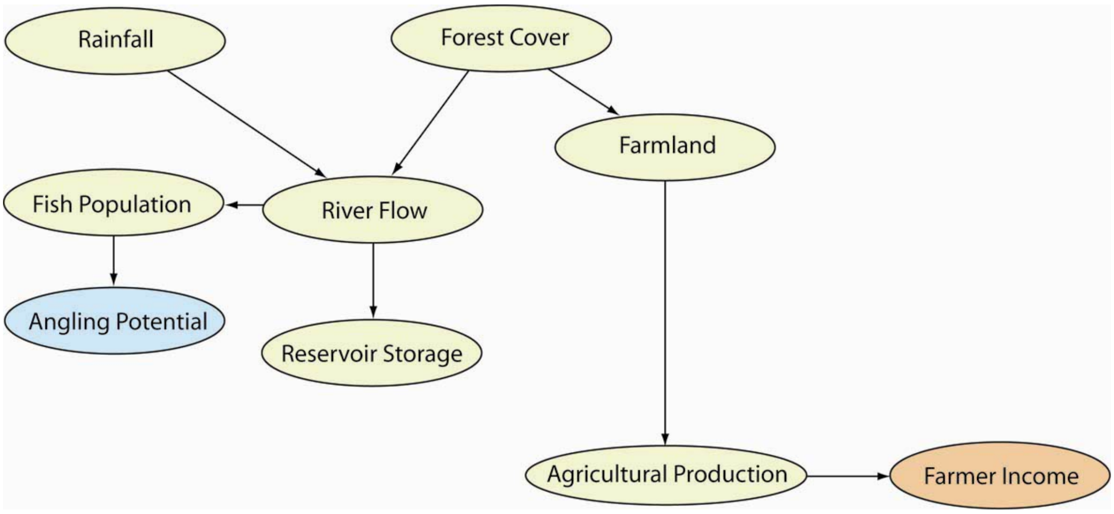


Figure 2.2 A simple Bayesian network

In this particular network there is only one action open to the manager and that is to change the amount of forest cover; the other variable directly influencing the outcome is rainfall, but this is outside the manager's control. Note that variables within the network can be of any type; in this case economic, physical and social factors are all included, indicated by brown, green and blue nodes respectively.

The node 'Forest Cover' is parent to two child nodes 'River Flow' and 'Farmland', while in its turn 'Farmland' is a parent to 'Agricultural Production', which is a parent to 'Farmer Income'. The way in which this network has been constructed means that any change in forest cover will have an effect on the area of land cultivation (farmland), which in turn will lead to a change in agricultural production and ultimately to the income of local farmers. But the impact of forest cover change is not restricted to farmer income. The network shows that changes in this parameter will also affect river flow and in turn the fish population and angling potential in the region, as well as flow to the local reservoir. Running the network will allow the user to evaluate the benefits and disadvantages of changing forest cover. Decreasing the area of forest may increase farmer income (economic advantage), but have an adverse affect on reservoir storage and angling potential (environmental and social drawbacks). On the other hand, increasing forest cover may lead to better angling and reservoir conditions, but at the expense of decreased farmer income. The network can be used to evaluate and weigh the pros and cons of each action, to help obtain a fair and balanced decision.

The ability to link diverse types of information in this way is a key characteristic of Bns and one which makes them particularly suited to the problems of integrated water management. The power to integrate information lies with the nature of the links, the behaviour of which is defined by the construction of conditional probability tables for each variable. These probability tables lie at the core of the network.

2.3.3 Conditional Probability Tables

The strength of a link between two nodes is expressed as a 'probabilistic dependency', which is quantified by a conditional probability table (CPT). Each variable within a network has an associated CPT, which expresses the probability of that node being in a particular state, given the states of its parents. For variables without parents, an unconditional distribution is defined; in other words it is the operator who decides on the state of the node. In this case the state selected will depend on the nature of the variable; it may be based on existing evidence of the state of the variable, or represent a scenario or potential action that may take place. Setting the states in this way is described as entering 'evidence'. Entering evidence in a variable as an observation, scenario or action, will result in a chain reaction of impacts on all variables linked to it. When a network is run with a new set of starting conditions the probability distributions reflecting the states of each linked variable is changed. This is illustrated in Figure 2.3 which represents a 'compiled' version of the network shown in Figure 2.2. In Figure 2.3 'monitors' have been opened in front of each variable to reveal their probability distributions.

In this example, for simplicity, the states for all nodes have been described in words (labels). The two controlling variables here are 'Forest Cover' and 'Rainfall'. For the particular case shown, evidence has been entered setting both variables in a 'high' state, indicated by the red bar, which highlights a 100% probability that the variable is in this state. The remaining windows show the probability distributions for the variables affected by the two controlling nodes. For example, in this instance we can see that with high rainfall and forest cover, there is a 39% probability that 'Farmer Income' will be good, but a 61% chance that it will be bad. At the same time the 'Angling Potential' of local rivers has an 88% of being good and only a 12% chance of being poor. Changing the evidence in either of the two controlling variables will result in a change in the probability distribution of those nodes to which they are linked. In this example reducing the area of forest will probably increase the income of farmers, but at the same time is likely to reduce fish populations and angling potential.

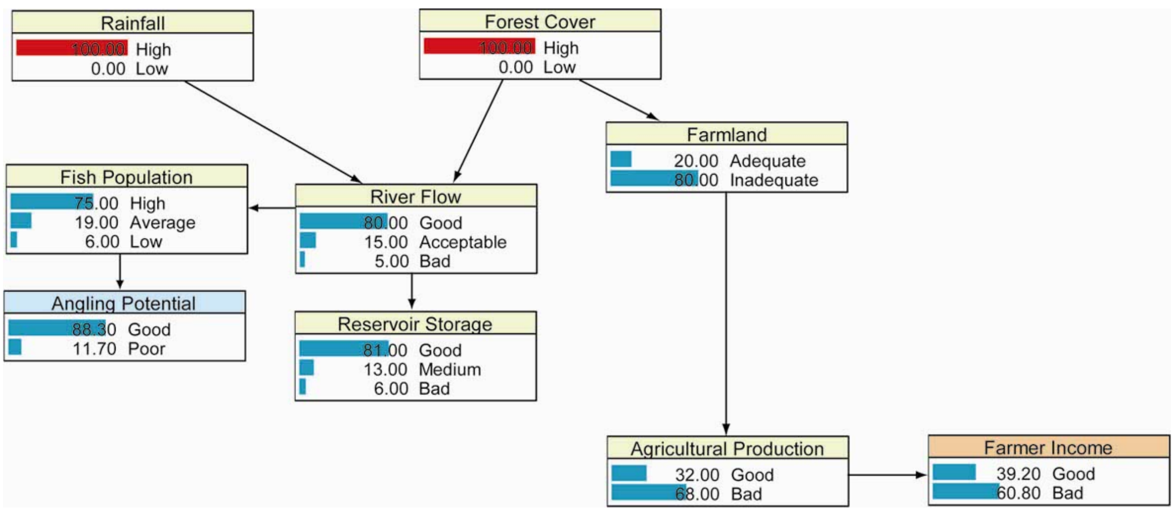


Figure 2.3 Compiled version of the simple Bayesian network shown in Figure 2.2

The presentation of results in the form of a probability distribution rather than single values is an integral feature of Bns. This explicit representation of the uncertainty attached to the prediction makes it an ideal tool for the field of water resource management because the complexity of the natural world means that it is rarely possible to predict the impact of a management option with any degree of certainty.

Probability distributions are calculated from the *prior* CPTs constructed for each variable and the evidence entered into the network. The CPTs constructed for each variable have the form shown in Table 2.1, which is for 'River Flow' and its child, 'Reservoir Storage' shown in Figure 2.2.

		River Flow		
		Good	Acceptable	Bad
Reservoir Storage	Good	*0.9*	*0.6*	*0*
	Medium	*0.1*	*0.3*	*0.1*
	Bad	*0*	*0.1*	*0.9*

How to read the Table. The second *row* gives the 3 states of 'River flow': Good, Acceptable and Bad. The second *column* gives the three states of 'Reservoir Storage': Good, Medium and Bad. If 'River Flow' is Good, the first column of figures shows there is a 10% (0.1) chance that 'Reservoir Storage' is Medium; if 'River Flow' is Acceptable, there is a 30% chance 'Reservoir Storage' is Medium, and if 'River Flow' is bad, there is a 90% chance 'Reservoir Storage' will be Bad. The other lines are interpreted in the same manner.

Table 2.1 Conditional Probability Table for the 'Reservoir Storage' variable in Figure 2.2

In this case a change in the state of river flow has a direct impact on reservoir storage. Based on the information available in the table it is possible to say that if we know the river is in an acceptable state (i.e. 100% probability), then there is 60% chance reservoir storage will be good, a 30% chance it will be medium, and only a 10% chance of being bad. Changing the probability distribution of the 'River Flow' node by changing the evidence of its parent nodes ('Forest Cover' and 'Rainfall' in Figure 2.2), will change the probability distribution of the 'Reservoir Storage' node.

The probability distribution for reservoir storage in Table 2.1 is based on the river having a 100% probability of being in any one of its three states. But of course we may not always be certain of the state of the river. For example, in Figure 2.3 the flow has an 80% chance of being good, a 15% chance of being acceptable and a 5% chance of being bad. Given this distribution the network uses the data in Table 2.1 to calculate the resulting state of the reservoir. A brief explanation of the way this is done is given in Appendix I.

The key to constructing a good network is to have the best available data with which to construct the CPTs, although the best data available may be imperfect and not wholly reliable. A particular strength of Bns is that they will accommodate any type of data, but of course the less reliable the information, the more uncertain will be the result and the wider the distribution of probabilities. Data may be obtained from three different sources, which are in decreasing order of confidence:

1. *Direct measurements:*
Extensive data sets, based on long term monitoring and/or extensive field measurements, provide the most reliable source of information. For example, Table 2.1 might have been based on a long term record of river and reservoir storage measurements, the longer the record the better.

1. *Output from models:*
Tables may also be constructed from the output of models. For example, a model might be used to provide information about the impact of changing forest cover on river flows, where direct measurements may not exist. The models can be of any

type, hydrological, economic, ecological or social, or those that link different disciplines. It should be stressed, however, that networks do not replace the models themselves, but simply accept output from models and converts it into a format suitable for inclusion within a CPT.

2. *Expert opinion:*
In some instances the data for a particular link will be limited, or even non-existent. In these cases it may be necessary to fall back on expert opinion. In Figure 2.3, changing the river flow will have an impact on the angling potential. But to quantify 'potential' is difficult, and relevant data is unlikely to exist. In this case an expert opinion, perhaps from someone representing the local angling clubs, may be used to help complete the CPT. With this type of subjective input the degree of uncertainty is likely to be greater than that obtained from measured data, but this uncertainty will be explicitly expressed in the output. The ability to use this type of input enables the Bn to overcome problems of data scarcity.

For more information on the statistical background of Bayesian networks readers are referred to Jensen (1996) and Neapolitan (2003). Some basic principles and properties of networks are given in Appendix I of these guidelines.

3. PRINCIPLES OF STAKEHOLDER ENGAGEMENT

3.1 Principles and Definitions

The Water Framework Directive encourages the active involvement of stakeholders in environmental decision making, whenever possible. For a decision to have any resonance with stakeholders, it is essential to encourage their active involvement. Furthermore, the earlier in the decision making process this involvement takes place, the better. This is not to say that every decision on every occasion will demand full public participation, but whenever possible participation should be encouraged. The degree of involvement cannot be prescribed, but will depend on local circumstances, the type of decision that is being made, and not least on the time and money available for the procedure.

This section of the Guidelines sets out some generic principles and working methods relevant to stakeholder involvement; further guidance on stakeholder engagement, specifically in the context of Bayesian network construction, is given in Section 4.

With respect to stakeholder engagement in general, these Guidelines endorse the following key principles:

1 The objectives of stakeholder engagement should be clearly stated

2 The nature / form of the engagement must be defined and developed in the local cultural / institutional / political context

3 As broad a range of interested parties / individuals as possible should be engaged

4 It must be transparent as to how stakeholders' views have been used in the development of the network and if they have not been used, why not

5 Experts must be willing to change / amend their views in the light of stakeholder input where possible

6 Sufficient time must be allowed to collect stakeholder views

7 Group, as opposed to individual, engagement can more effectively generate new ideas or thinking. However, use of both methods can optimise access to different people

8 Stakeholder engagement must be seen to impact on the decision to be made

9 There should be evidence of enhanced stakeholder understanding of the issue being discussed – i.e. social learning

The term *stakeholder* is taken here to mean interested parties (i.e. those who have a stake in an issue). These interests may be regulatory, business / financial, or individual. Interests may arise through participation in organised bodies or groups or as individual citizens. Stakeholders are considered, therefore, in their broadest sense, and it is important to remember that even if people choose not to declare their 'stake' they still have a 'right–to-know' if their interests may be affected by a decision.

The term *engagement* is used deliberately, contrasting with concepts of consultation. Consultation implies asking people for their views once a proposal or plan has been formulated in draft. Here it would imply consulting people once the Bn had been drafted in full and applied in the decision context.

Engagement on the other hand implies that people are asked in advance of any draft plan, and that their input has a real impact on the drafting. In this case it is engagement in the construction of the Bn. Engagement implies a process of active discussion or deliberation, not passive consultation whereby people are merely asked for their views – often by letter or sometimes through some form of questionnaire.

These principles draw on generic good practice in stakeholder and public engagement in environmental decision-making but are developed here in the specific context of the application of Bns to river basin management. Examples are drawn from the practice of stakeholder engagement during the MERIT project.

One important issue to note is that effective engagement in the Bn is likely to require further engagement in other elements of the decision, not least as stakeholders' awareness and knowledge of the particular river basin management issue is raised.

3.2 Stakeholder Engagement Process

Stakeholder engagement might include (Figure 3.1):

- Assistance in identification of key issues (framing or scoping the problem)
- Identification of relevant data to feed into the Bn
- Oversight of the development of the Bn (to ensure specialist inputs)
- Involvement in assessment and evaluation of the tool's outputs and agreement on these in relation to the specific decision context (to act as QA mechanism and reach consensus on the outputs)
- Broader consultation on the outputs of the tool (as an information provision exercise, but also to receive comments)

As Figure 3.1 suggests stakeholder engagement should be considered as a process, not a single activity at a point in time. Stakeholder engagement is essential to the development of the Bn, indeed it should be integral to it.

Engagement does not merely inform the development of the Bn, rather it is part of the development process itself. The stages of the development process are discussed in detail in section 4.

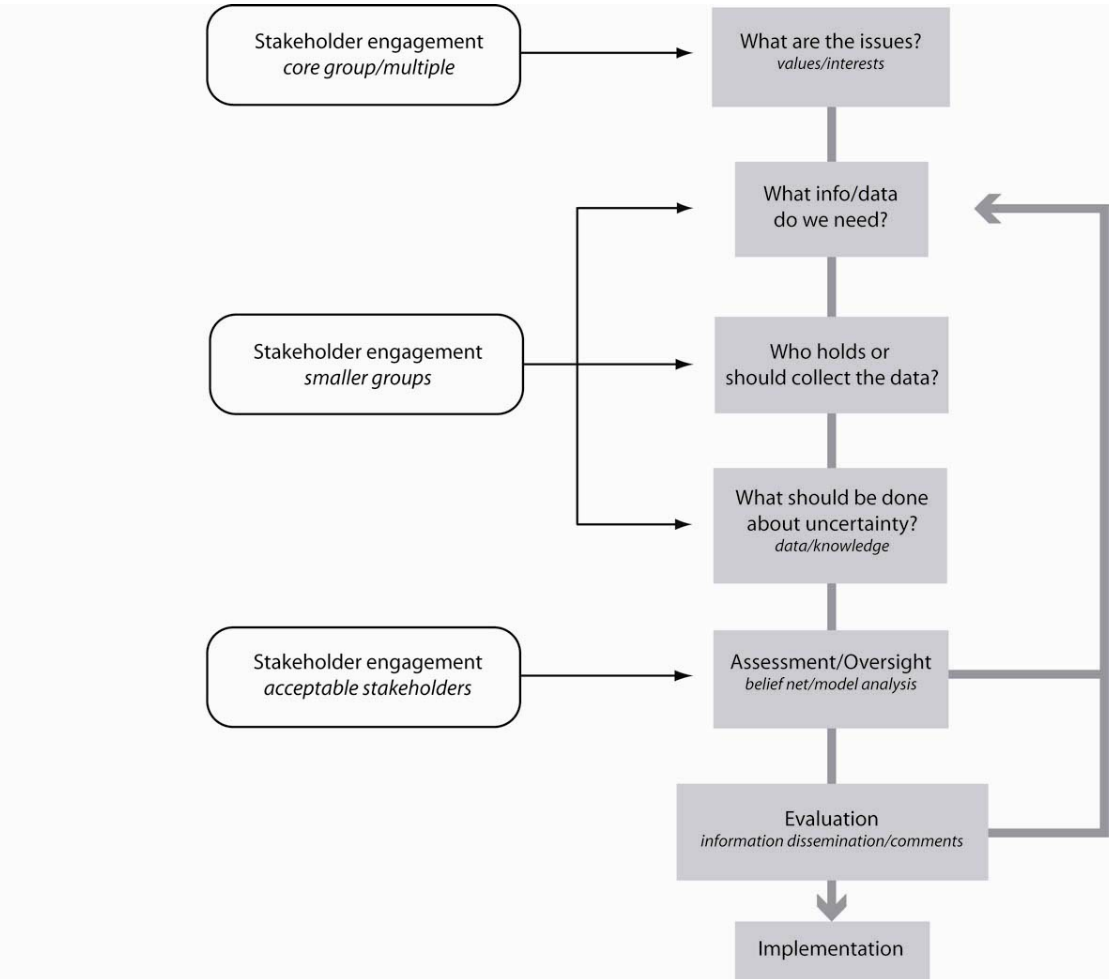


Figure 3.1 – Stakeholder Engagement Process (Petts, 2004[7])

3.3 Objectives and Rules of the Game

Any engagement process has to be 'fit for purpose'. While it is possible to identify elements and principles of good practice, any context or site-specific process of engagement requires a design that is relevant to that context. It is not possible to replicate approaches from country to country, nor indeed is it always possible to replicate a process within a country in different areas or decision contexts.

Within MERIT the country-specific approaches were not only different because the river basin contexts were different (i.e. groundwater quality; hydropower development; water resource management), but because the cultural contexts in terms of experience and expectations of engagement were different.

For example, in Denmark there is a long history of public engagement in all types of decisions; people expect to participate and to be asked to participate. The style of government and governance is consensual and open.

[7] Petts. J. (2004) Barriers to participation and deliberation in risk decisions: evidence from waste management. *Journal of Risk Research*, 7(2), 115-133

This contrasts with Italy and Spain, which have tended to a more closed style of governance with decisions tending to be 'top down' and limited to consultation of relevant interests. However, even in these countries expectations are changing, with legislative development in response not least to the UNECE's[8] Aarhus Convention (*Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters).*

In the UK, the water resource context in which the Bn was tested was already subject to a process of limited stakeholder consultation – i.e. under the Environment Agency's 'Catchment Abstraction Management Scheme' (CAMS). Therefore, MERIT had to fit with this process and was limited in the extent to which new stakeholders could be involved. There was also limited opportunity for the Bn to influence a decision, compared for example with the case in Denmark.

At the start it is important that all those involved – the builders of the Bn as well as stakeholders - are clear as to the objectives of engagement and how the engagement process will work. The objectives should include recognition that stakeholders can meaningfully contribute to the development and building of the Bn as well as comment on its output.

The objectives should clearly state the limitations of what can be achieved, not least in terms of the extent to which a real decision can be influenced. People are less likely to be willing to engage (particularly over any length of time) if they do not believe that their inputs will have any influence. They are less likely to engage if they consider that the issue has no relevance to them (hence the focus on engagement with those having an interest).

Relevant objectives could include, for example:

- To identify people's concerns and interests
- To resolve conflicting interests
- To gather local information and information based on practical experience
- To test expert knowledge
- To increase the transparency of the Bn development process
- To enhance social learning – i.e. people's understanding of the issue
- To reach a consensus view on the most appropriate Bn for the decision process

Note a consensus does not imply agreement, but simply that majority views are addressed and minority views at least noted.

3.4 Representation

Defining interests

A key question underpinning concerns amongst experts and decision-makers about engagement is whether, when views are gathered, they can be considered to be representative of the majority of people who are not directly engaged. Questions are raised for example about whether activist environmental groups represent the public. Sometimes, experts question whether limited engagement with a small group of people can provide a genuine set of representative views.

[8] UNECE: United Nations Economic Commission for Europe. Details of the Aarhus Convention can be found on: http://www.unece.org/

It is important to draw a distinction between '*representing interests*' and '*being representative of interests*' or 'representation' versus 'representativeness'.

Representation is akin to sampling. However, it raises questions about the faithfulness of the representatives to the views and knowledge of those people who are represented. Representativeness extends to identifying people whose knowledge and views are likely to be representative of a certain set of interests – e.g. of farmers; anglers; water companies. People are not asked to represent the views of these interests (this would imply that they would need to ask all other farmers or anglers for example about their views). But they are considered to have experiences, knowledge and concerns similar to the larger group.

The broader the range of interests involved, the more likely that the Bn will be inclusive in terms of knowledge and concerns (Box 3.1).

Box 3.1 Defining interests

Interests can be broadly considered in terms of:

- **Institutional** – e.g. local/municipal authorities
- **Professional** – e.g. groundwater expert from a research organisation; industrial concern; conservation group
- **Local –** e.g. local farmer or resident

Of course professionals are ultimately individual citizens as well as experts. It is the context in which they are asked to contribute that is relevant. Importantly, the selection of representatives must be as inclusive and open as is possible and would be expected in the context. This will enhance the legitimacy of the engagement process and lead to a higher level of trust in its outcomes.

Within MERIT one interesting issue emerged in Spain. Here individual citizens were not recruited, but in practice a number of the professionals who were actively involved were also citizens (farmers) in the area. Their contributions were enriched by personal experience as farmers while they focused on their technical knowledge. As local farmers, they would potentially be perceived to have interests and views representative of other farmers in the area, although they were engaged as experts. They had roots in the local community.

Identifying relevant interests

One mechanism to identify potential interests is to list categories of water users and institutions in the area, and then to define their different types of interest and whether they are likely to be for, or against, the management initiative (as was done in Denmark). Box 3.2 lists examples of interests in river basin management issues.

It is important to have a range of views and not to ignore or side-line the views of those who may be against any initiative. Such views will always arise at some point and deliberately avoiding the input of known potential opponents will only lead to possible further antagonism and delays in progress with the decision.

In both Denmark and Spain in the MERIT project the potential interests of stakeholders were initially explored by an inclusive invitation to participate through gathering of information about experience and problems (Box 3.3).This approach not only provides for an initial scoping of concerns that need to be considered, but also for the identification of any relevant interests and / or individuals who have been missed and who should be asked to participate.

In Denmark the initial stakeholder meetings also provided an opportunity to agree the objectives, means and process of engagement.

Box 3.2 Example Interests in River Basin Management

Institutional and Professional	Local and Community
• Authority/municipal – water protection, land use planning, conservation • Water supply companies/corporations • Corporate industrial and farming water users • Industrial and farming unions • Environmental groups • Conservation groups • Wildlife trusts • Fishing organisations • Water research institutes and scientific bodies	• Water users – homeowners/residents, individual farmers • Those with private water supplies – residents, farmers • Those potentially at risk from flooding – homeowners, farmers • Private landowners • Recreation interests – anglers, boating enthusiasts • Members of local conservation and wildlife projects and groups

Box 3.3 Interests, Identification and Definition

In *Denmark* a letter was sent to all professional stakeholder organisations found to have a potential or even marginal interest in groundwater protection in the area with an invitation to attend an initial one-day workshop. Following this, a public meeting was held in the local community with invitations distributed to more than 1100 local households. At both meetings stakeholders were asked to present issues and problems they regarded to be important to groundwater protection

In *Spain* a stakeholder questionnaire was sent to 22 expert, institutional and professional stakeholders (public administration; water managers; water users; ecologists; agricultural workers' unions and scientific experts) to identify the capacity to manage water resources, the use of the resources, technical, legal, environmental and economic problems

Spatial dynamics

Spatial dynamics can be an important consideration where issues such as river basin management are involved. There may be regional or even national interests relevant to a specific river basin, as well as local interests arising from defined responsibilities for the water resources in a particular area, or from people being resident in the area. Spatial dynamics can extend an engagement process, but nevertheless need to be considered. Occasionally it might be the case that there are trans-boundary interests and it may be necessary to address the interests of those in a neighbouring country.

3.5 Methods

There is no single appropriate method of engagement replicable in all circumstances. Only the *principles* of engagement are replicable. These are:

- Transparency of process and of decision
- Open to all interests
- Participants representative of the broad range of likely interests
- Clarity of purpose
- Opportunity for dialogue and discussion

The construction of a Bn in the river basin management context has a number of distinguishing features that promote the adoption of dialogue-based methods (workshops, focus or discussion groups) as opposed to information gathering processes: e.g. one-to-one interviews; questionnaires; etc. (Box 3. 4)

Box 3.4 Characteristics of Bns that favour dialogue-based methods

- Scientific and technical complexity (including potentially technical or expert language)
- Inherent uncertainty that needs to be characterised and understood
- Data availability problems – data both published and unpublished may need to be revealed
- Divergent expert / scientific opinions
- The iterative nature of Bn construction to combine knowledge and data into a single model
- The underlying divergent values, interests and experience that may need to be reconciled and balanced
- The need to make trade-offs

However, there are also inherent difficulties with Bns that make them appear less amenable to discussion – not least their technical nature, which may make it difficult to motivate stakeholders to get involved. As revealed in Denmark there can be concerns about the “Bn being a toy for researchers”.

If the decision issue is sufficiently important to people, *and* it is clear that stakeholder views will contribute to that decision (i.e. decision-makers minds are not already made up), then engagement will be easier and more valuable.

Box 3.5 Example Stakeholder Priorities in Denmark

Farmers Association
- Genetically modified crops seen as solution for pesticide agriculture
- Need to understand cost of increased price of water for consumers
- Conflict between afforestation and other land uses
- Want the best value for money in terms of clean drinking water

Private Waterworks Association
- Water levy difficult to explain
- Pesticide pollution from private gardens a problem
- Evident conflict between willingness to pay and the desire for clean water

Citizens
- No perceived problems with nitrate pollution in the area
- Restrictions should be placed on industry not agriculture
- Need to maintain the river to avoid flooding
- More afforestation required
- Private households should stop using pesticides in gardens

The need to make trade-offs will be important. In Denmark the initial professional stakeholder workshop and the local community public meeting revealed (or confirmed) a set of divergent priorities and interests that had to be considered in constructing the Bn (Box 3.5), but also provided the framework for data and technical inputs to the Bn.

The full nature and basis of such concerns can most effectively be revealed through dialogue-based processes. These also assist in raising the understanding of divergent interests across different groups.

Mixing stakeholder interests

Methods that can potentially engage different interests – institutional, professional and local, such as workshops – need to be considered as to the benefits and difficulties (Box 3.6).

In Denmark a 2-stream parallel involvement process – professional / expert and citizen / public – was run. The two stakeholder groups worked separately but came together in a final joint meeting. Information from each was available to the other throughout. As Figure 3.1 suggests the process should ensure the input of relevant people at appropriate times. Thus a large workshop or public meeting (for 10s of people) with multiple stakeholders is a good mechanism for scoping or framing the issues that need to be addressed. They work well in the more qualitative stages of the Bn development, whereas a small group, perhaps less than 20 people, can be more effective where technical data need to be discussed in detail – the more quantitative stages.

Smaller working groups can be formed from larger groups. Indeed one way to enhance the acceptance of the outputs of a small group (which might be considered not to be suitably inclusive) is for the group members to be identified and selected by those in a large group so that it is agreed that the participants will usefully represent all interests.

Box 3.6 Benefits and Challenges of Mixed Workshop Discussions

Benefits	Challenges
• Enhances understanding of all interests and concerns • Brings groups and individuals together who would not normally have an opportunity for discussion • Provides for trade-offs to be agreed • Opens expert knowledge to lay or local knowledge • Builds relationships that can be ongoing	• Differential time needs for understanding an issue – for example, local residents may need longer to understand the nature of the issue than people from a water company. • Differential conditions of trust – local interests may have low inherent trust in industry and / or authorities • Technical language and professional jargon used by some participants may be difficult for others to understand

Non-dialogue components – multiple methods

Engagement guidelines stress the importance of **multiple methods** being used – i.e. combining methods of information provision, consultation and dialogue – to enhance overall effectiveness and to ensure that the maximum number of interested people have an opportunity to comment.

In the context of Bns it is apparent that one-to-one methods (interviews, telephone discussions, meetings) are required at times during the development process – not least during the quantitative stages of Bn development where data suggested or provided by stakeholders may need to be checked and understood in more detail. All of the MERIT case studies involved one-to-one methods.

Information provision through leaflets, newsletters and similar information modes can be an important mechanism to ensure that people not directly engaged in discussion processes are aware of the work being conducted and the decision context. This is important for local communities, ensures transparency and provides opportunities for non-participants to be involved if they wish.

Facilitation

Dialogue-based processes such as workshops and focus groups benefit from independent facilitation. This:

- Enhances the openness of the process of collecting opinions and data as the person facilitating the discussion is understood to have no specific interest
- Allows the technical specialists responsible for constructing the Bn the freedom to do their job without having to manage discussion and debate
- Provides an effective mechanism for different views to be heard and for dissenting voices not to dominate
- Provides an independent mechanism by which the engagement process can be effectively managed to meet its objectives, while at the same time ensuring that the concerns of stakeholders, information needs and opportunities for discussion are adequately provided for. This is a difficult balance to achieve.

Where the facilitator has some understanding of the issue being discussed, this enhances their ability to move debate and discussion forward and to ensure that key points of discussion are allowed appropriate time. Although the use of a professional facilitator is encouraged, it should be noted that it all adds to the time and cost of the process.

Stakeholder engagement is resource intensive. This is the primary reason why it is sometimes not undertaken effectively – i.e. an attempt to reduce or avoid the costs. However, short term savings can lead to long-term costs through later opposition to proposals because of a failure to involve people early and build in, or at least transparently address, stakeholder interests and concerns.

Stakeholder engagement requires time. In the context of Bn development the technical complexity of the data and the uncertainty inherent in the tools means that iteration of concerns and data is essential. River basin management issues often involve conflicts of concerns and priorities – for example, about water resource uses. Engagement will reveal these conflicts, but if time is cut short and people do not feel that sufficient attention has been paid to an issue, trust in the process will be diminished.

For lay stakeholders in particular it is important to allow time for them to gain an understanding of unfamiliar issues – this may require multiple meetings. People cannot effectively join in discussion until they feel confident that they understand what is being discussed.

It is unlikely that any effective Bn development work can be done with any stakeholders in less than 2 dialogue-based meetings. Box 3.7 summarises the timeframe for the case study in Denmark.

The problem with long timeframes, such as in Denmark, is that the stakeholder groups can lose momentum when nothing seems to be happening. It can be difficult to maintain interest and even to bring people back to the table. It is important to keep people informed of what is happening and why.

Box 3.7 Process timeframe in Denmark

In Denmark the engagement process involved:

- Initial workshop and public meeting
- Citizens' group meetings with facilitator
- Three workshops with professional stakeholders
- Individual meetings with citizens and professional stakeholders
- Final joint professional and citizens meeting

The period required for these actions was from November 2002 to February 2004.

The costs of organising and facilitating meetings need to be fully covered. Simple organisational issues such as providing refreshments are important. Also it is likely that stakeholders, particularly local citizens, will need expenses (e.g. travel). It is common to pay a small honorarium to people to recognise the value of the time they give to the process. In the UK an honorarium of approximately £100 (€150) for 4 half-day / evening meetings (or equivalent), is common for members of the public.

3.6 Transparency

Any engagement process should be transparent and open from inception, through operation, to communication of its outcome. Consensus and dissenting views must be acknowledged in the outcome.

Within the Bn, the process of construction must be sufficiently transparent, open to challenge, and to the input of alternative assumptions if agreed to be appropriate by stakeholders. One effective mechanism to ensure this happens is to record in a diary-type format how the pilot network and then the quantitative data are incorporated into the Bn. The logic behind the network elements, the reasoning behind the data inclusion, and adjustments to probability assumptions must be clear to people.

The significant advantage of a Bn is that all of this can be displayed visually. However, maintaining a record of the reasons for the changes (whether stakeholder or expert generated) is an important element of transparency.

Adequate opportunities must be provided for the uncertainty inherent in factual information to be recognized, explained and discussed, and knowledge deficiencies to be openly acknowledged.

Opportunities must be provided for expert knowledge and information to be challenged, and any claims to be tested through discussion. Where it becomes apparent that information or knowledge is missing, or insufficient, resources should be made available for further work to be undertaken. The minutes of all stakeholder meetings must be produced and agreed with the attendees. The format of minutes should be agreed at the start (e.g. whether individuals should be identifiable).

An important characteristic of the Bn is that the tool can be used interactively for uncertainty assessment and communication. However, experience is that it is important to apply a kind of 'protocol' for the Bn construction in order to explain to all those involved (experts, team members, users and stakeholders) which inputs are required at different stages (Box 3.8).

Box 3.8 Protocol for development of a Bayesian Network

1. Define the context
2. Identify factors, actions and indicators
3. Build pilot network
4. Collect data
5. Define states
6. Construct conditional probability tables, and
7. Collect feedback from stakeholders

Such a protocol makes it easier for all players to 'play the game'.

3.7 Summary – Why do it?

Experience in the Bn context (and indeed from engagement in a whole range of environmental issues) is that:

- People are interested
- People can engage with highly technical issues if given sufficient time
- People can understand risk issues and are willing to take a pragmatic approach to risk balancing
- Networks are not a familiar approach to people – they are a way of thinking that requires a little exercise
- People contribute valuable local knowledge and experience

An analysis of the strengths, weaknesses, opportunities and threats of the stakeholder engagement process is presented in Box 3.9.

3.8 Stakeholder engagement plan

To ensure that all those engaged understand the rules of the game experience indicates that a Stakeholder Engagement Plan is necessary. The Danish experience suggests that the plan should include the items listed in Box 3.10.

Box 3. 9 Summary SWOT Analysis of Stakeholder Engagement (Petts & Leach 2000[9])	
Strengths Engagement can… • Bring out technical knowledge from the public and others • Use local knowledge not known to the authority • Encourage diverse perspectives (and so identify issues not thought of) • Allow people to understand the system better • Use people's passion and enthusiasm • Enable a better evaluation of the issues	**Weakness** Engagement can be weakened by… • A lack of resources (time, money, staff) • An inadequate legal framework • A lack of awareness / experience of participation • Difficulties in gaining access to information • A lack of technical support for the public • Limited consideration of the results of participation • Not enough engagement
Opportunities Engagement offers the opportunity to… • Build trust and capacity • Improve the environment, build a community and avoid wasting resources • Empower people by starting a dialogue and improving openness • Expand the limits of understanding (working together to solve problems) • Prevent conflicts by early involvement of the public • Save time in the overall decision process by reduction of opposition	**Threats** Engagement processes can be threatened if… • People think that the process is a formality (that minds are already made up) • A vocal minority dominates public meetings • Not enough time is allowed to make a decision or discuss the proposals • The long-term implications are not understood (e.g. if planning gain wins over the long-term interests) • Technical and expert submissions are not of good quality and do not cover all the issues

Box 3.10 Example Stakeholder Engagement Plan

- Develop a common understanding of issues and concerns
- Define goals and objectives
- Define team roles and responsibilities
- Identify facilitators
- Identify potential stakeholders
- Evaluate interests and responsibilities
- Create terms of reference for all groups
- Form discussion groups and plan meetings
- Agree timetable of events and rules of conduct with stakeholders
- Identify and describe resources required
- Detail initiatives for general information provision to the general public and other stakeholders

[9] Petts, J & Leach, B. (2000) *Evaluating Methods for Public Participation: Literature Review*. R& D technical Report E135. Bristol, Environment Agency

4. BUILDING THE MODEL: A GUIDE TO NETWORK CONSTRUCTION

4.1 The route to a successful network

There is no single way to construct a successful network. A host of factors will influence the type of approach; the cultural, social and economic setting, the number and nature of the stakeholders involved, and the objectives of the exercise will all play a part, not to mention the constraints of time and cost. The experience of MERIT has shown that networks can be constructed using different routes to accommodate local conditions. Nevertheless, it should be stressed that whatever procedure is used, a management strategy is far more likely to be successful if stakeholders are encouraged to become involved with the design and construction of the network at an early stage. Such involvement creates a culture of transparency, will generate a sense of 'ownership' toward decisions that are made, and provide the environment within which the resolution of conflict is more likely to be achieved.

This section deals with the stages of building a Bayesian network and the way in which stakeholders can be involved at each step. The process of construction has been broken down into 7 stages:

Stage 1: Define the problem and select appropriate approach
Stage 2: Identify variables, actions and indicators
Stage 3: Design pilot network
Stage 4: Collect data from all available sources
Stage 5: Define the states for all variables
Stage 6: Construct conditional probability tables
Stage 7: Check, collect feedback and evaluate network

Each *stage* is described in two *sections*:

The first section explains *network construction and design*. Headings for these sections are in **orange** font. The second section of each stage looks at the way in which *stakeholders* can contribute to the process of construction. Here the headings are in **blue**.

By and large, the technical steps involved in network design and construction are the same regardless of the problem being addressed. It is not practical to skip any of these steps, though the order in which they are done can be changed. In contrast, the extent to which stakeholders contribute to the construction process can range from practically nothing, to complete involvement to the point where they significantly influence the final decision. In these guidelines there is an underlying presumption that as much stakeholder involvement as possible is encouraged at all stages, while recognising that time and financial considerations may limit what is actually possible.

Examples from the four MERIT case studies are used throughout to illustrate and reinforce particular aspects of the process. A brief summary of each case study is presented in Section 5; full working versions of each Bn, with explanatory descriptions, are included on the attached CD.

4.2 STAGE1: Define the problem and select appropriate approach

The first stage of Bayesian network construction is illustrated in Figure 4.1. Yellow boxes refer to technical aspects, those in blue to stakeholder involvement. The diamond shaped box represents a decision that needs to be made during the course of this first stage. The steps for Stage 1 are numbered thus: step 1.1 (orange font) refers to the first step of network construction; step 1.1 (blue font) is the first step in stakeholder involvement.

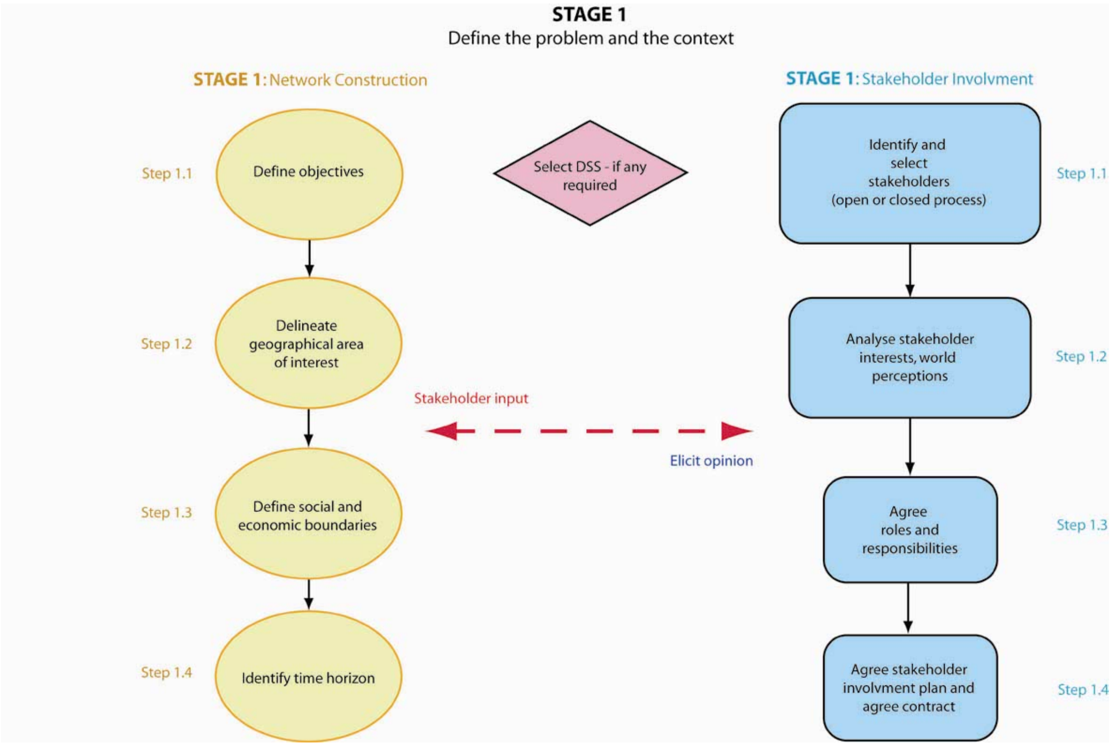


Figure 4.1 STAGE 1: Define the problem and select appropriate approach

It should be stressed at the outset that these guidelines are not meant to be prescriptive, but simply to provide a general framework which can be modified to suit your particular situation. The precise procedure will differ from region to region. But whatever approach is taken, the principles of engagement do not change; the process must be transparent, be representative, open to all interests, be clear, and offer the opportunity for dialogue and discussion.

Stage 1: Network construction

Before network construction begins it is important to clarify precisely what you want to achieve, who you wish to involve in the process, and how you wish to achieve your objectives. Stage 1 of network construction involves the following:

- *Define the problem and objectives*
- *Define physical boundaries*
- *Define social and economic boundaries*
- *Identify a time scale for the network*

Once this initial stage has been completed, you will be in a position to decide whether or not you need to use a decision support tool. It may be that the problem involves a small number of variables and so does not require the use of a Bn or any other type of decision support system. Or it may be that time or financial constraints preclude the use of a Bn.

1.1 Define the problem

Before any decision process begins it is essential to have clear objectives. What do you hope to achieve through your management intervention in the system?
The problem can be broken down into two simple questions.

1. What variables am I trying to impact?
2. What actions are available to me to make this impact?

The network will be used to assess the impact of different actions, or combinations of actions, on the variables you are trying to influence. In some cases, instead of actions you may wish to analyse the impacts of different scenarios – a 'what if' type of problem.

The variables you are trying to influence can be termed 'indicator' or 'objective' variables, which will be used to judge the success or otherwise of a management option. They may be of any type e.g. economic (farmers' income), physical (water quality).

The larger the number of action and objective variables chosen, the more complex will be the network, and the more difficult it will be to reach a decision; so at this stage careful thought should be given to how wide-reaching the decision process should be. The scope of the decision process should be discussed and agreed with stakeholder groups at an early stage (Box 4.1).

1.2 Define the physical boundaries

Be clear about the geographical area you want to cover. This may be a catchment, but it can be any type of region such as an aquifer unit or a water demand area. But whatever area is chosen, you must be consistent throughout the network. For example, if a catchment is selected it is self evident that the data and calculations for all variables have to based on the catchment unit (Box 4.2).

1.3 Define the social and economic boundaries

It is important to recognise that the social and economic limits of your network may not be confined within the physical boundaries that have been set. Account should be taken of external social, economic, legal, technical or cultural factors that may have a significant impact within the area. Variables such as the level of subsidies set by the EU and national governments, the market price of crops and the cost of transport, are all examples of external factors that may influence the economy of the area covered by the network. Some external variables used in the Spanish case are shown in Figure 4.2. Here legal, economic and technical factors, external to the area, are included because they have an impact on the water resources of the region.

Box 4.1 Define the problem and objectives

In the Danish ‘Compensation Network’ the objective was clear: to investigate the impact of different levels of compensation payment to farmers to encourage them to reduce the amount of pesticide application in the catchment. The problem in this area is that pesticide is contaminating a major aquifer used for drinking water by the city of Copenhagen, so there is an understandable desire to stop, or at least reduce, the level of contamination.

Within the ‘Compensation’ variable there are 7 levels of compensation ranging from 0 to 4400 Dkr ha^{-1}, which represent the potential actions open to the water manager, Copenhagen Energy. This variable is a ‘Scenario Defining Node’ (SDN).

Copenhagen Energy needs to know what impact each level of compensation will have, not only on groundwater quality, but on a whole range of other factors that need to be taken into consideration. This is the reality of integrated management. Six ‘indicator’ variables representing different interests in the catchment have been identified to be of prime importance. These are:

1. Groundwater Quality
2. Biodiversity
3. Incidence of Biological abnormalities
4. Recreation value of resulting land use
5. Hunting potential of resulting land use
6. Income from crop production

In this network we thus have a set of 6 ‘indicators’ whose states are to be tracked in response to a range of potential compensation payments.

Box 4.2 Select geographical area and time scale

The networks completed for the MERIT project each have different types of <u>geographical area</u>. In the Danish example the selected area was a hydrological region, the Havelse catchment, to the north-west of Copenhagen. A *catchment* was also selected as the basis for the UK network, this was the Loddon catchment in south east England. However, in the Spanish case the study area was an *aquifer unit*, the East La Mancha Aquifer, which supplies most of the water used for irrigation in the region. Finally, there have been two networks developed in Italy, both using different types of geographical area. The first network is based on the *area of irrigated land* potentially available for development along the lower reaches of the Vomano River. The second, which examines the environmental performance of different water management policies, has the capacity to use different geographical areas depending on the requirements of the user. These include areas based on (a) catchments (b) fish habitats and (c) National Parks or Special Areas of Conservation (SaCs).

So far as the <u>time period</u> is concerned, all the networks were based on one year, except the irrigation network for the Vomano River. This network is run and updated on a daily basis, because it is linked to a reservoir operation model which calculates flow daily.

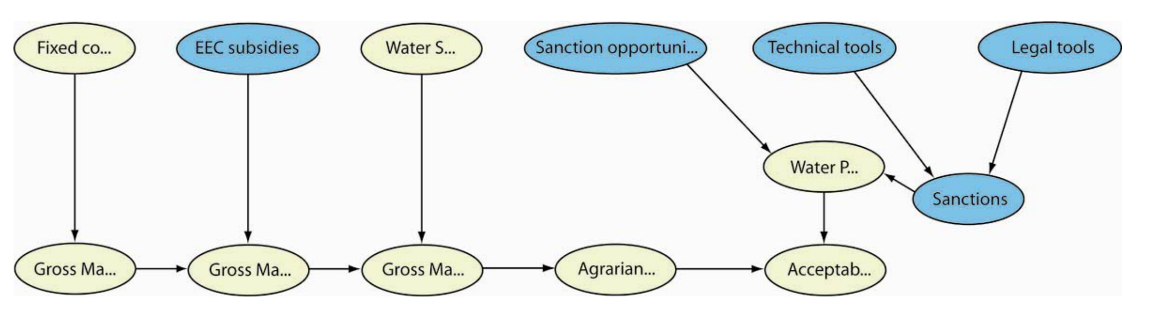


Figure 4.2 Example from the Spanish network showing external economic, technical and legal factors in blue

1.4 Define the timescale

Decide on the time scale you wish to use. For water resource issues the most convenient time period is a hydrological, or perhaps an agricultural year. A longer time period may be used, but remember in a complex environmental system it becomes difficult to predict more than 3 years in advance; so remember the further you look into the future the more uncertainty there will be (Cain, 2001).

Decision to use an Environmental Decision Support System (EDSS) - or not

During the course of Stage 1 the methodology to be used for the decision process will need to be selected. It may be that the problem is straightforward, involving a small number of options, and affecting a restricted number of variables. In this case an examination of available data may be sufficient to make an informed decision, so that an Environmental Decision Support System (EDSS) is not really required. However, in most cases environmental systems are too complex to deal with in this way, and an EDSS is needed to help the decision making process.

There are many types of EDSS and many 'generators' for these systems, of which Bayesian networks are but one. Others include Influence Diagrams, Decision Trees, Multi Criteria Analysis and Mathematical Models. But before deciding which type of EDSS is the most appropriate for the problem being considered, it is useful to review what you want from the system.

1. **Do you want the internal working of the environmental system to be represented?** You need to decide to what extent it is necessary to represent the internal working of the environmental system. In some cases actions and causes are self explanatory and no explicit representation of the process is needed. In the Danish 'compensation' case it is fairly clear that a contract drawn up, based on a particular compensation scheme, will affect farm economics; no explanation of the process is really needed. On the other hand the way in which different compensation schemes may impact on the quality of deep groundwater does require some explanation. An outsider cannot be expected to automatically make this link. When weighing up the degree of complexity to include, it is important to achieve a good balance between having too much and too little detail. If there is too little detail there may be insufficient evidence to justify decisions, too much makes the network complex and difficult to understand.

2. **Is it important to communicate the reasons for your decision to stakeholders?** Where it is necessary to justify a decision, the best means of

communicating the reasoning to stakeholders, and possibly the wider public, has to be an important consideration.

3. **Is it necessary to incorporate the views and opinions of stakeholders within the DSS?** It is generally important to represent within the EDDS the opinion, views, and perhaps information, from stakeholders. The system must, therefore, be a transparent representation of all important factors and their interdependence.

4. **Is uncertainty associated with your decision? If so how important is it to represent this uncertainty within the EDSS?** Because you will be dealing with complex environmental systems it is inevitable that some uncertainty will be attached to the data used by the EDSS, and to the decisions based on that data. It is necessary to be aware of the risks involved with the decision being made, and the consequences of failure.

If you decide that (a) some representation of the internal operation of the system is required, (b) justification for the decision is needed, (c) stakeholders need to be represented within the EDSS and (d) uncertainty will be associated with any decisions, or any one of these, then a Bayesian network offers a good way to proceed with the decision making process.

At this point it should be stressed that if it is to be done effectively Bayesian networks, or any other type of EDSS, require input of time and money. Organising and running stakeholder meetings cannot be rushed or conducted without incurring costs; likewise collecting and analysing the data for the EDSS is time consuming and will incur additional expense. You are, therefore, advised to estimate the cost of using an EDSS before you start, to help decide whether the advantages to be gained outweigh the expense and time involved.

Stage 1: Stakeholder involvement

At Stage 1 of construction, the following actions to involve stakeholders are needed (Figure 4.1):

- *Identify and select stakeholders*
- *Identify the views, interests and concerns of stakeholders*
- *Agree roles and responsibilities*
- *Agree 'stakeholder involvement plan' contract*

1.1 Identify and select stakeholders

What is a stakeholder?
A stakeholder can be either an individual, or organisation, that has a vested interest in the outcome of the decision being made. Types of organisation may include:

- water providers (e.g. water companies)
- local government authorities (e.g. local councils)
- potential polluters (e.g. industry, farmers)
- leisure groups (e.g. sailing clubs, angling societies)
- environmental / conservation groups (e.g. Greenpeace, Friends of the Earth, Wildlife Trusts)
- water users (e.g. farmers' organisations)

Individuals can also be stakeholders. These may have a general interest in environmental matters or be someone with a specific concern about the effect of water management on their situation. When selecting members for a stakeholder group it is important to strike a balance, with all points of view being fairly represented. Before making your selection, it is worth listing all potential stakeholders by category, as above. Once this has been done the selection of individuals to represent the different interests can be based on a number of criteria. Stakeholders should:

- Command authority within the organisation they represent
- Ideally, live or work in the study area
- Possess local knowledge
- Be good at assimilating and assessing technical information
- Be able to attend all meetings or workshops
- Be able to work to a tight timetable
- Have a genuine interest and concern about the problem(s) in hand

Making initial contact.
Selection of the right group of stakeholders is critical to the success of any decision making process. Unfortunately there is no prescribed selection procedure; it depends very much on the nature of the decision to be made, and local circumstances. Ideally selection should begin as soon as the need for a decision becomes apparent. In some cases the opinion and knowledge of stakeholders could even be used to help frame the nature of the decision itself. Initial contact can be made in a number of ways:

- Distribution of leaflets by post, hand or public display, describing the issues and inviting interest
- Telephone contact to invite interest
- Calling a public meeting
- Contacting specific organisations or individuals directly to invite participation

Methods used by the MERIT partners are described in Box 4.3, along with examples of stakeholder groups for two of the networks.

One note of caution about initial contacts: if a public meeting is called and large numbers attend, it may prove difficult during the meeting to select the right individuals to represent the public (citizens') point of view. Often meetings are dominated by a few vociferous and articulate people whose views may not necessarily represent those of the majority. Public meetings are an effective way to inform, but not necessarily the best way to select members of a stakeholder group. There is also the time and cost of organising a public meeting to be considered.

Organisation and size of stakeholder groups
The size of stakeholder groups should be sufficiently large to encompass all shades of opinion, yet not so large that it becomes cumbersome and expensive to manage. In MERIT, the groups ranged from 6 to 9 in number, a size that proved to be effective and easy to manage. It is recommended that group sizes should be kept below 10 if at all possible.

In some cases you may wish to split the groups. In the Danish case, stakeholders were divided into a 'citizens' group representing the opinion of the non-expert, and a 'professional' group including organisations and individuals deeply involved in the problems of water management. Splitting the groups ensured that the professionals

Box 4.3 Initial contact with and selection of stakeholders

The UK 'Water Demand Network'. Over 400 leaflets describing the objectives of the 'Catchment Abstraction Management' exercise were posted with an invitation for people to raise issues or express an interest in the exercise. The leaflet campaign received a 10% response rate, which is good for this type of initiative. Stakeholders were selected on the basis of the responses received and included all the major institutions and organizations directly affected by the exercise.

The Spanish 'La Mancha Aquifer' network. A total of 22 questionnaires were sent to targeted individuals and organizations (public administrators, water managers, water users, ecologists, agricultural workers' unions, scientific experts). The aim was to assess the capacity to manage water resources, and identify the technical, legal, environmental, and economic difficulties being faced in the region. Such a selective survey was justified on the basis that members of the MERIT team had many years experience of water management in the region and were able to identify those stakeholders who needed to be involved in the network construction.

The Danish 'Compensation' and 'Flooding' networks. The starting point for stakeholder identification was to categorize the list of water users, potential polluters and authorities in the local area. Letters of invitation to a workshop were sent to all professional stakeholder groups. Non-professional (citizen) stakeholders were selected following a public meeting advertised in the press and by leaflet distribution to more than 1100 households; the meeting was attended by 100 people.

The Italian 'Reservoir' and 'Environmental performance' networks. Initial contact of stakeholders for the 'Reservoir' network was made by telephone, followed up by face to face meetings. Stakeholders for the 'Environmental network' were all expert groups of ecologists and environmentalists already known to the network constructor.

Examples of stakeholder groups

UK 'water demand' network	**Danish 'compensation' network** ('Professional Group')
South East Water (*water provider*)	County Organisation (*local government*)
National Farmers' Union (*water user*)	Greater Copenhagen Authority (*government*)
Wildlife Trust (*conservation*)	Municipalities (*local government*)
Local Council (*local government*)	Farmers Union (*water users / polluters*)
Fisheries Consultative Committee (*leisure*)	Danish Industry (*water user / polluter*)
Recreation management Service (*leisure*)	Wildlife trusts (*conservation*)
	Private landowners (*water users*)

were not able to use their expertise to dominate meetings and suppress alternative views emerging from the citizen representatives.

Decisions are generally made by regulatory bodies that are often responsible for the organisation of stakeholder consultations. However, if the consultation is organised by another organisation, it is important that representatives of regulatory bodies who have an influence on the final decision, but who are not directly affected by it, are included. A group without the inclusion of such an authority will clearly be far less influential.

1.2 Identify the views, interests and concerns of stakeholders

All stakeholders, either representatives of organisations or individual citizens, will have their own self - interests and prejudices. In most cases little can be done about these biases, but when building a stakeholder group they should be recognised and taken into account. Every effort must be made to ensure that all points of view are fairly represented in the group; this is particularly important in cases where conflict is likely.

It must also be recognised that the world perceptions of stakeholders are not easily changed, even in the face of what you may consider to be overwhelming evidence. For instance, in the Havelse catchment many years of monitoring, analysis and modelling has strongly indicated that pollution of deep aquifers is caused by current agricultural practices. However, despite such evidence local farmers continue to maintain that this is not the case.

The different perceptions of all stakeholders should be given consideration, even if it is initially not seen by the ‘experts’ to be a valid concern. The Danish network again provides a good example. Although the original problem in the Havelse catchment was considered to be pollution of groundwater resources, a number of the citizens’ group expressed concern about flooding in the lower part of the catchment. So strong was the feeling that a second network was developed to address this issue.

In the UK case there was a different outcome. Stakeholders expressed a worry about the effects of water abstraction licensing on the flow of a canal, however, sufficient scientific evidence was available to convince the group that the issue was not important, and it was dropped.

1.3 Identify stakeholder roles and responsibilities; role of facilitator

Stakeholder roles and responsibilities

The roles and responsibilities of each stakeholder should be listed. Particular note should be made of any ongoing or planned stakeholder activities in the catchment. The information can best be obtained through a questionnaire, though verbal contact may be sufficient. This exercise will help you (a) to understand the position of each stakeholder with respect to the decision to be made, (b) identify any legal or political restrictions there may be and (c) help pinpoint possible sources of information for inclusion in the network. Roles and responsibilities of selected stakeholders from the MERIT case studies are listed in Table 4.1.

Case study	Authority	Role	Activities	Legal
Denmark	*Copenhagen Energy*	Provide safe water supply to Copenhagen	Modelling, geological mapping, monitoring, license negotiations	Legal requirement to supply water
Italy	*Enel*	Provide Hydro Electric power	Operation of reservoirs	Total water rights and control over river flow
UK	*South East Water*	Private water company; water supply provider	Abstraction, water supply	Legal requirement to supply water
Spain	*Irrigation Users*	Represents farmers	Provide water for irrigation use	

Table 4.1 Roles, responsibilities (legal requirements) and activities of a selection of MERIT stakeholders

Facilitator
It is recommended that, if possible, stakeholder groups should be chaired by independent facilitators who have the experience to control and direct meetings, particularly when conflict between stakeholders is likely. Although this adds to the cost, the inclusion of a facilitator can mean the difference between success and failure. Two of the MERIT cases used facilitators. In the UK network an experienced person, not associated with the CAMS[10] procedure or with the area concerned, was appointed by the Environment Agency to chair all meetings. In Denmark the facilitator for the citizens' group, and some of the professional group meetings, was a person from the local joint municipality Agenda-21 centre.

The presence of a good facilitator ensures:

- Independence and assurance to the stakeholders that the meetings are not being manipulated for the benefit of the decision makers
- That meetings are efficiently chaired, well focused and likely to achieve the set objectives
- That potential conflicts can be diffused without detriment to any of the group members
- All stakeholders have the opportunity to present their points of view, and that meetings are not dominated by a small number of vociferous members
- The presence of an independent sounding board for the decision maker

The potential drawbacks of experienced facilitators are that they are firstly expensive, and secondly may not be available at times convenient to you or the stakeholders. Where the decision to be made is local and low key then the need for a professional person is probably not justified, but in cases where the decision has a far reaching effect, is controversial and likely to be placed under intense scrutiny, a facilitator is essential.

1.4 Agree 'stakeholder involvement plan', and contract

One of the main reasons for failing to successfully involve stakeholders in the decision making process is that at the outset the expectations, responsibilities and objectives of the exercise are not made sufficiently clear. Experience has demonstrated that *clear rules of the game* are very important. These need to be prepared as a written agreement during this first stage of network development. Such a *stakeholder involvement plan* should include the following:

1. An agreed common understanding of the problems or issues being faced; 'moving the goal posts' at a later stage in the process will undermine stakeholder confidence
2. A clear statement of the nature of the public participation expected. This will make clear to what extent the stakeholders will be able to mould, change and influence the decision being made. If the exercise is purely intended to inform, rather than involve the stakeholder, this should be made clear
3. A definition of the role and responsibilities of all the participants, including the decision maker, individual stakeholders and facilitator. Responsibilities will include attendance of meetings, distribution of minutes etc. It is particularly important to ensure that stakeholders will be able and willing to attend all the planned meetings

[10] CAMS: Catchment Abstraction Management procedure, adopted by the Environment Agency in the UK for the control of licensed abstractions from catchments

4. A comprehensive list of all stakeholders
5. An evaluation of each stakeholders' interest, responsibility and activity in the region
6. The formation of separate working groups, if necessary. This may be the case where conflict between stakeholders is intense. Placing people in conflict together without initial preparation may simply polarise opinion and result in a stalemate
7. Appointment of a facilitator to ensure that meetings deliver their objectives
8. Creation of 'mission statements' for all groups, so that everyone has a clear idea of what they are trying to achieve
9. A schedule of future meetings and milestones
10. A description of the way in which the stakeholder involvement process will be funded and resourced e.g. cost of venues and facilitator, expenses, printing etc.
11. A description of the way information will be made available, if at all, to the general public

By the end of Stage 1 you will have:

- Identified objectives, area of study and timescale
- Identified and selected your stakeholder group(s)
- Decided to use a Bayesian network as a decision support tool
- Agreed and drawn up a stakeholder involvement plan

4.3 STAGE 2: Identify variables, actions and indicators

Stage 2: Network construction

The second stage of network construction covers the identification of all variables relevant to the problem in hand (Figure 4.2). Selection of these variables should be done in consultation with the stakeholders as described in the Stage 2 Stakeholder Involvement section.

So far as the technical aspects of construction are concerned Stage 2 requires the following actions:

- *List all variables relevant to the problem*
- *Identify key indicators*
- *Identify potential actions (interventions) and / or scenarios*
- *Identify the available data sources*

Each variable represented in the system is different, but to make the network easier to organise Cain (2001) suggests they be grouped into categories that describe their function. He suggests the 6 categories shown in Table 4.2. These categories provide a useful means to view the logic and structure of a network without getting too involved with the detail. We shall be referring to these categories throughout the guidelines where a description of the general function of a variable is required.

2.1 List all variables relevant to the problem

The overall objectives and scope of the decision making process have been set in Stage 1. In Stage 2 we move on to compile a list of all the variables considered relevant to the problem identified (i.e. all the categories shown in Table 4.2). An initial

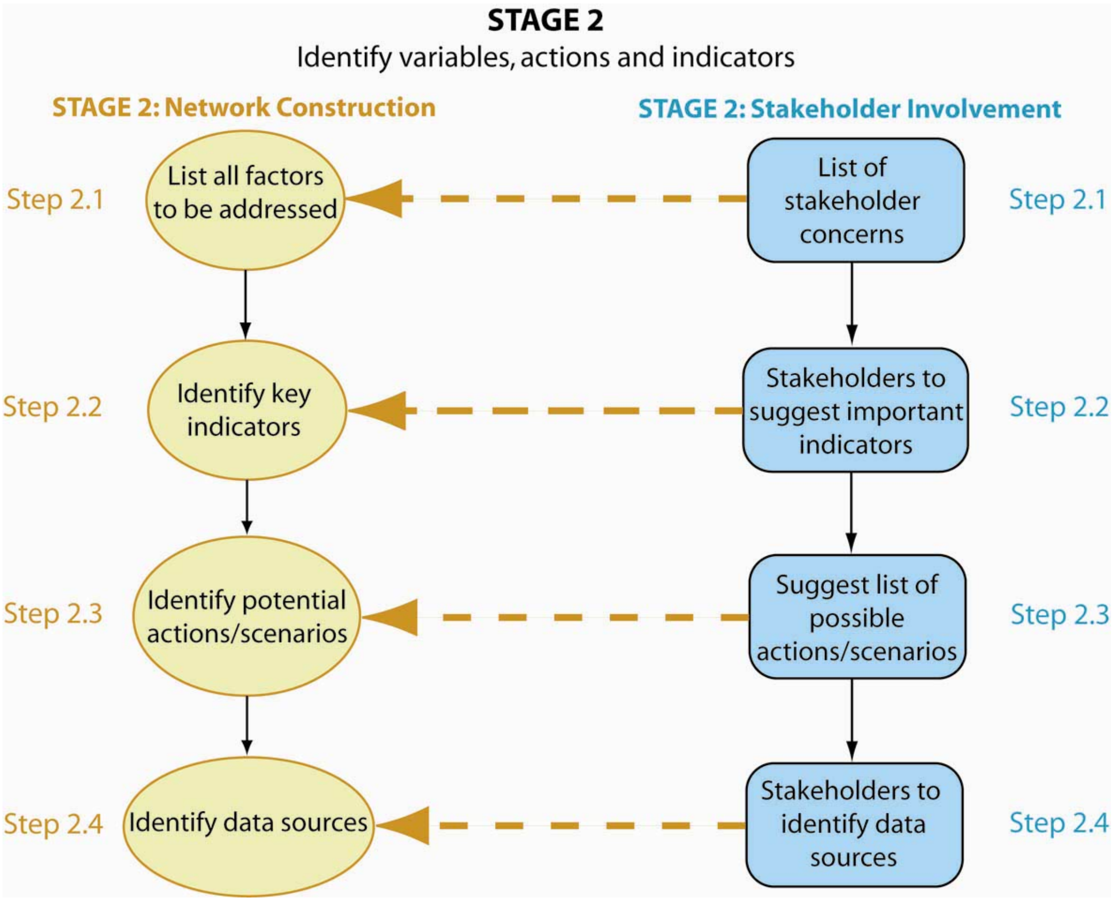


Figure 4.2 STAGE 2: Identification of all variables and data sources

selection can be made by the decision maker and / or other experts familiar with the problem and the area. However, a complete list can only be finalised after consultation with the stakeholder groups (steps 2.1 - 2.4, stakeholder section).

Selecting variables is all about trying to capture ideas. But ideas can be expressed in different ways and it is important to capture the concept you are trying to represent in a way that is easily quantifiable and readily understood by the non-expert. Capturing ideas and representing them in this way requires a particular way of thinking and it can sometimes take a while before the network constructor is able to produce a list of variables that are appropriate.

For example, how do you represent the effect of a management strategy on a variable such as 'regional groundwater levels', something that varies in both space and time? One option might be to use the water level records from representative boreholes set against target levels for each location; another might be to use the number of dry wells in the area as an indicator of groundwater level; or perhaps a measure of the water available for irrigation could be used as a surrogate indicator of the state of groundwater resources. Selecting the most appropriate way to represent different concepts is an important step that should not be rushed.

Sometimes a concept might be represented as a single variable, but at other times it might require more than one. Where possible, and without sacrificing transparency, it is advisable to keep the number of variables to a minimum. There is no hard and fast

Category	Description	Examples from MERIT
Objectives	These are variables that you hope to affect through different management strategies or scenarios. These will include the 'indicator' variables used to judge the success or otherwise of your management choice	• Deep groundwater quality (DK) • River outflow (ES) • Metered water consumption (UK) • Water supply deficit (IT)
Interventions	These are the things you need to implement to achieve the objectives. It can be a physical action (e.g. increase forest cover), an economic intervention (e.g. change water price) or any other type of action that has an impact on the objectives	• Compensation payments (DK) • Number of new houses (UK) • Capital subsidies (ES) • Irrigation district enlargement (IT)
Intermediate Factors	Factors that link objectives and interventions	• Pesticide load - between compensation and deep groundwater quality (DK) • Grey Water savings - between Number of new houses and New house consumption (UK) • Available capital - between Capital subsidies and Irrigation efficiency (ES) • Potential evaporation - between solar radiation and supply deficit (IT)
Controlling factors	Factors that influence the system in some way but which cannot be controlled e.g. Climatic conditions	• Sand / Clay soils (DK) • Number of existing houses (UK) • CAP subsidies (ES) • Solar radiation (IT)
Implementation factors	Factors that directly affect whether an intervention might be successful such as available funding for the action, the support of local communities or availability of resources.	• Legal tools to enable sanctions (ES) • Acceptability (of water plans) (ES)
Additional impacts	Factors that are changed as a result of interventions but that do not affect anything else in the environmental system. If they are not considered important or relevant they can be excluded.	Not used

Table 4.2 Categories of variables adapted from Cain (2001). The examples can be viewed in the networks on the CD accompanying the guidelines.

rule but in some cases it may be better to combine two or three variables into one, provided the combined variable is capable of being easily understood and quantified. Then again, there are other occasions when it is better to sub-divide a single variable into more than one.

The Danish network provides an example of combining variables. Here the variable 'sand / clay' is used to indicate the degree of pesticide pollution likely to take place in the shallow groundwater aquifer. In this case the percentage of sand and clay in the soil is being used as a surrogate to represent a complex process which involves far more than this simple description. In reality factors such as the thickness of clay /

sand, the slope, the presence of cracks or fissures, the soil moisture status, rainfall amount and intensity and so forth, will all play a part in controlling how much pesticide finds its way down to the water table.

However, given the objectives of the Danish network, it is simply not necessary to include these variables, even assuming the data were available. Of course if the objective was specifically to investigate pesticide transport to the shallow water table, then inclusion of variables relevant to the transport process would be essential and need to be included in the network.

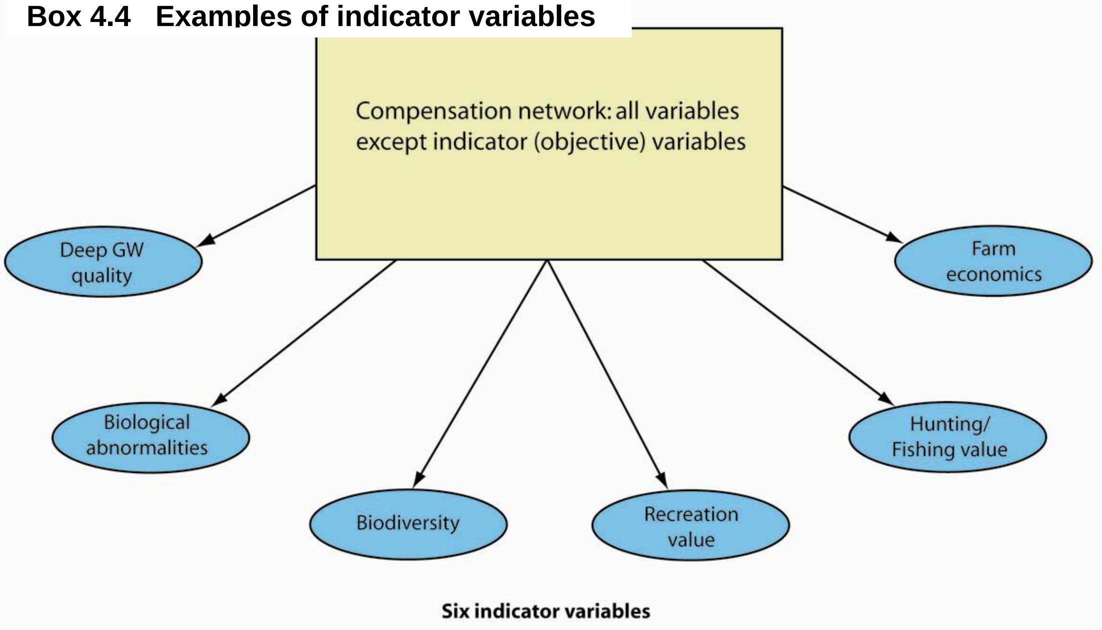
In contrast, the UK network offers an example of a case where a single variable has been replaced by several. In this network the effect of different water demand measures on household water consumption is examined. But rather than simply having one variable ‘Household savings’ to indicate the overall amount of water saved, there are instead five, each representing different components of household water use (toilet flushing, garden watering etc.). This decision was made because a subdivision of water use in the home was considered to be relevant in the context of the overall network.

2.2 Identify key indicators

Indicators are a key element of the network because it is the states of these variables that are used to identify the success or acceptability of a management strategy or scenario. Consequently they need to accurately represent the aspect of the system you are trying to preserve or protect.

In many cases the indicators selected will tend to act against each other, particularly when economic and environmental variables are being evaluated. In the Danish compensation network if the state of the indicator ‘deep groundwater quality’ improves, the likelihood is that the state of a second indicator ‘farm economics’ will worsen; this is because water quality improves at the expense of reduced pesticide application, which in turn affects crop production and ultimately farm income. Examples of indicator variables selected from the Danish network are shown in Box 4.4.

Box 4.4 Examples of indicator variables



When selecting indicators care should be taken to ensure that they are capable of being easily quantified. An important objective might be the necessity to maintain an 'attractive landscape' for social recreation. However, it is difficult to quantify this type of subjective variable. Instead something more measurable such as river levels, vegetation cover, vegetation type etc should be considered instead.

2.3 Identify potential actions and / or scenarios

Among the variables you need to identify are the management actions that could potentially be implemented to influence the state of the indicator variables.

Alternatively a series of different scenarios to represent changing conditions in the future may need to be identified. This selection should be done and ideally agreed with the different stakeholders involved in the process (step 2.3, stakeholder section). Examples of selected actions and scenarios for the MERIT project are shown in Box 4.5. These implementation variables do not generally have parents, but instead are instantiated (i.e. the states are set) by the user.

2.4 Identify data sources

The collection of data for the network does not begin until later in the process, but at this early stage it is useful to identify potential sources; knowing the type, quality and amount of data to hand may influence the selection of variables and design of the network. It is also important to establish whether stakeholders possess, or are aware of, data sets that are not in the public domain, but which can be used (Stage 2, stakeholder involvement).

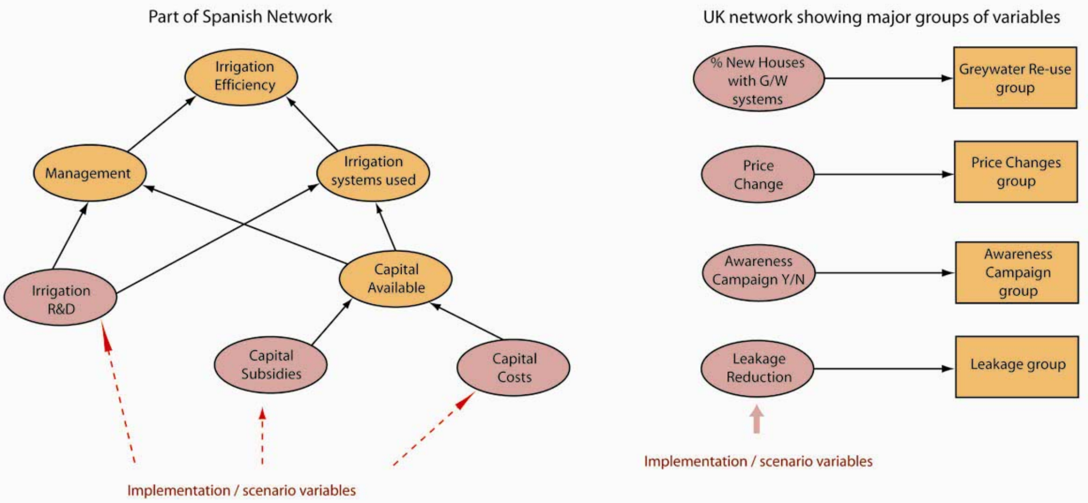
Stage 2: Stakeholder involvement

At this stage you should explain to the stakeholder group(s) about Bayesian networks and the way they will be used in the decision making process. The degree of technical complexity you choose to convey will depend on the composition of the stakeholder group. In the case of a largely professional group of experts a more detailed description of the methodology can be given; where the group is made up largely of non-experts the explanation can be simplified. However, be aware not to underestimate the ability of ordinary members of the public to understand networks. One of the criticisms directed at the Danish case study was that the 'specialists underestimated the ability of 'citizens' (non-experts) to comprehend the significance of the models'. People who have taken the time and effort to become involved in your project are likely to be intelligent and motivated, so avoid any temptation to patronise.

There is no right or wrong way to explain the methodology, but it is probably best to teach little and often throughout the process rather than trying to do everything at once. At this point it is probably enough to run through the basic concepts without going into too much detail. It is important for stakeholders to feel comfortable with the networks because any lack of understanding or misapprehensions will inevitably raise doubts in the stakeholders' minds. Also bear in mind that it takes time to become familiar with new concepts, even for the most intelligent person, so do not try to rush the learning process and be prepared to repeat explanations if they are not at first fully understood. Trying to speed the process up is a natural reaction on the part of decision makers.

A final point concerns the organisation of stakeholder meetings from this stage of the process onward. It is important that stakeholders are sent a detailed

Box 4.5
Examples of implementation / scenario setting from the UK and Spanish networks



agenda and objectives of planned meetings well in advance in addition to other information that may be relevant. In this way valuable time can be saved and the stakeholder feels that he / she is being kept fully informed of what is happening in good time.

2.1 List stakeholder issues and concerns

Before starting to design your network you should first ask all stakeholders to identify the problem or problems that need to be dealt with and the issues and concerns they have with respect to the problem. Inevitably these will be linked to their special areas of interest, so an angler will be concerned about the effect on fish populations, while for the water company representative the impact on water abstraction will be more of a problem. Provided the group has been selected to be representative, the list of issues should provide a comprehensive record of the most important concerns in the region.

This information should be incorporated in your list of network variables, even if you may not consider them to be important. The justification of each variable can be discussed at a later stage when you begin to rationalize the network. For an example the types of issues and problems raised by stakeholders in the UK network see Box 4.6.

Box 4.6

Issues raised by the UK network stakeholder group relating to the Loddon catchment

Stakeholder group	Issues raised
Wildlife Trust	• Needs of ecologically sensitive wetlands should be taken in account • Better identification of potential Sites of Special Scientific Interest
BVRMC (Recreational)	• Impact on Basingstoke canal should be investigated • Problem of water quality not addressed • The review is a ‘snapshot’ and should be repeated periodically
Farmers’ Union	• Leakage from the canal is a major factor to be considered • Suggestion that farmers should store water for summer use is unfair • Abstraction estimates from the catchment were thought to be too low
Local council	• The impact of increased housing targets need to be considered • Climate change not taken into account • Heavy localized abstraction anticipated in the near future
Water Companies	• Reminded the group they are legally obliged to supply water • Water companies operate on a 25 year time frame • A series of ‘what if’ scenarios are required to plan for extreme events
Fisheries Committee	• There is a need to protect the upper reaches of streams and rivers • Abstraction points should be located downstream • There is an effluent discharge problem both in rural and urban areas • Water quality in Basingstoke must be monitored if planned expansion takes place

2.2 Stakeholders to suggest list of indicators

Stakeholders should next be asked to consider what factors they would like to see influenced or controlled in the context of the problem being addressed. One stakeholder may wish to see the ecological health of a wetland maintained, while another is more concerned with ensuring a good crop yield. These will become the indicator or objective variables in your network and be used to judge the success or otherwise of any particular action or scenario. Care should be taken with their selection. Avoid vague objectives that are difficult to quantify or for which there is very little data. To provide the network with a clear focus and make it easier to use and understand, it is also advisable keep the number of indicators as small as possible.

Stakeholders should be asked to make clear what they wish to see preserved or changed in the area. Some of the suggestions will be readily acceptable, but others will be either too vague or there will be insufficient data available to merit their inclusion. It is important in discussion with the stakeholder group to decide whether unsuitable variables can be modified or expressed in a different way to allow inclusion. If too many indicators are suggested you may need to draw up a priority list and exclude those agreed to be less important.

You can use the selection of indicators as a means to concentrate minds on the key elements of the problem and exclude issues that may not be relevant. It is tempting to make the network ‘all encompassing’, but when this happens it often becomes ill-defined and of little practical use.

2.3 Stakeholders to suggest list of possible actions or scenarios

Once a list of indicators has been obtained you should next ask the stakeholders to suggest solutions to achieve these objectives, or in the event of a scenario-setting network, what type of scenarios should be simulated. Solutions will be in the form of actions that can be taken. Expert groups will be in a good position to put forward solutions that are both technical and practical. But the contribution of non-experts should not be dismissed. On the contrary, because this group is less constrained by traditional thinking, they may sometimes offer different perspectives and new insights that would not emerge through more conventional reasoning. Moreover, non-experts will have local knowledge that could be relevant to the viability of certain actions, ruling out some but including others.

Stakeholders should also be asked to list the range of ‘what if’ scenarios they consider to be the most important and would like to see investigated. Once again you should encourage the contribution of non-experts in the group.

The resulting actions and scenarios will become potential intervention or scenario-defining nodes for your network.

2.4 Identification of data sources

The reliability of your network will ultimately depend on the quantity and quality of data available to construct the conditional probability tables. Published data can be accessed by means of a literature search, but other information of a specifically local nature may be more difficult to obtain. This is where stakeholders can be very useful, either by being aware of local data sources, or perhaps possessing it themselves.

Representatives of organisations such as Water Companies, Wildlife Trusts, and Farmers’ Unions are liable to have suitable data for the network. But bear in mind that these data are likely to have been monitored over many years and be of considerable value to the organisation. You should establish at an early stage whether or not you can have access to the data. Sometimes it may be available, but only at a cost; on other occasions it may be confidential and cannot be released to third parties. Not being able to obtain key data sets may force you to reconsider the objectives, form and structure of your network. During the construction of the UK ‘water demand’ network, the privately owned water company freely gave access to information provided to OFWAT, the regulatory water body, but was unable to provide a more detailed breakdown for commercial reasons. This restriction influenced the way in which the network was structured.

A further source of information is output from models. If available, these may be in the hands of specialists who need to be consulted as ‘experts’. Sometimes the models can be run by the network constructor, thus saving the need for further consultation. An example of the use of model output is the Italian ‘reservoir’ network, which uses this type of data to calculate plant growth, potential evaporation and water distribution (Figure 4.24). A model is also used to generate data as input to the Italian ‘environment performance’ model (Figure 4.22b). The variable ‘WUA’ – the Weighted Useable Area – is calculated using models to simulate unmeasured flow and habitat conditions.

4.4 STAGE 3: Design pilot network

Stage 3: Network construction

Once the objectives have been identified, a stakeholder group or groups have been formed, and a list of all variables relevant to the problem identified and agreed, it is time to construct the first version of the network.

The steps making up stage 3 are shown in figure 4.3.

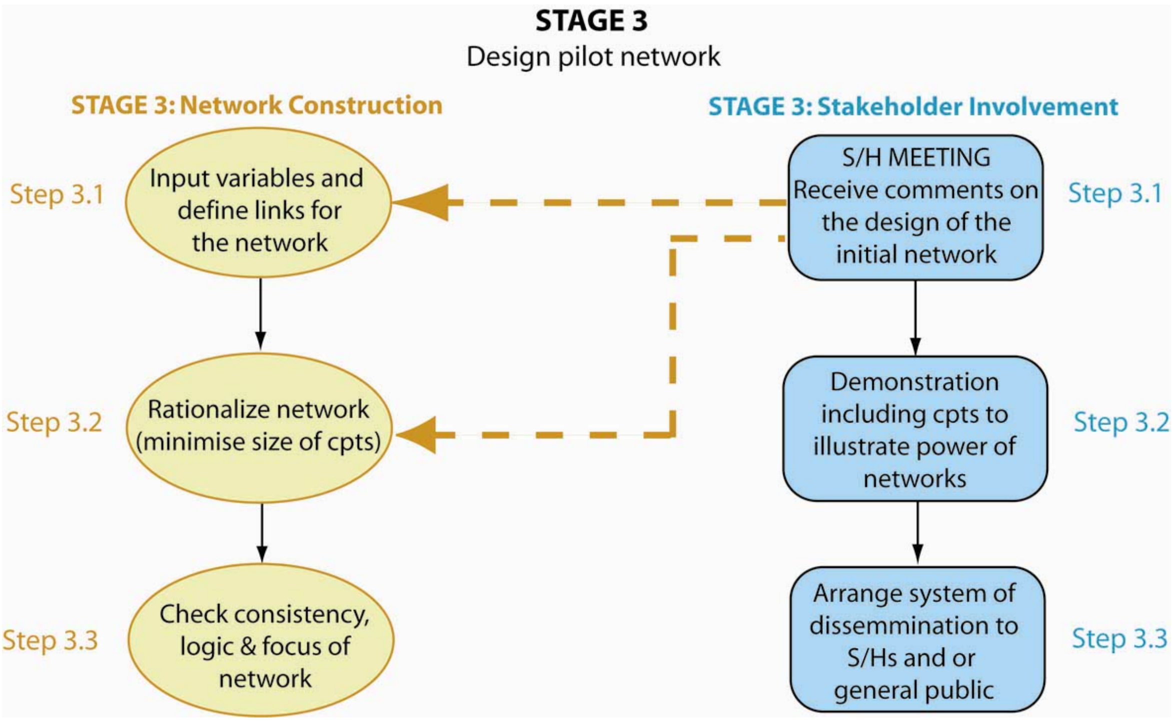


Figure 4.3 STAGE 3: Construction of pilot network

3.1 Input links and variables

You are now in a position to design and construct the network. The network should be a logical representation of the issues and problems agreed in Stage 1, using the variables selected in Stage 2. Begin by specifying the *indicator* variables. These are the factors you are attempting to influence through different actions or scenarios. Next put in the *action* variables representing the different management options that are available to achieve the changes you wish to implement. Alternatively, if you wish to examine the response of the system to changing conditions in the future, you may want to add scenario defining nodes; or of course you can have both types.

We next need to add the *intermediate* variables (Table 4.2). These represent the processes responsible for the cause and effect between the action and indicator variables. Together with the stakeholders, you enter the nodes and links you believe best conceptualises the system. This is very much a process of trial and error. How quickly and effectively the network is constructed will depend upon the complexity of the problem, and how well it has been defined.

Do not expect to complete a final network in one session. Be prepared to need several attempts. In the process of construction you will probably find problems and issues that were not previously identified in Stages 1 and 2. Your knowledge and understanding will be fully tested as you attempt to conceptualise the system, and it is more than likely that your original concepts will be modified. Without exception, all the networks developed in MERIT have been through several iterations; in the Spanish case 21 versions were produced during development (Table 4.3).

Network	No of versions
Denmark: Compensation network	6
Denmark: Flooding network	3
UK: Water Demand network	4
Italy: Irrigation network	3
Spain: La Mancha aquifer network	21

Table 4.3 The number of versions produced for each MERIT network

Some of these changes were made in response to changing perceptions of the problem as a result of stakeholder input; some resulted from a realization that the system could be captured in a different way; others stemmed from the necessity to limit the size and scope of the network.

The capacity of networks to force designers to think through the logic of actions and effects is one of the strengths of Bns. Even at this stage, without any states or numbers in the conditional probability tables, networks provide a useful tool to help provide a good conceptualisation of the system. At this stage they could be termed a 'Discussion Support System'.

3.2 Rationalising the network

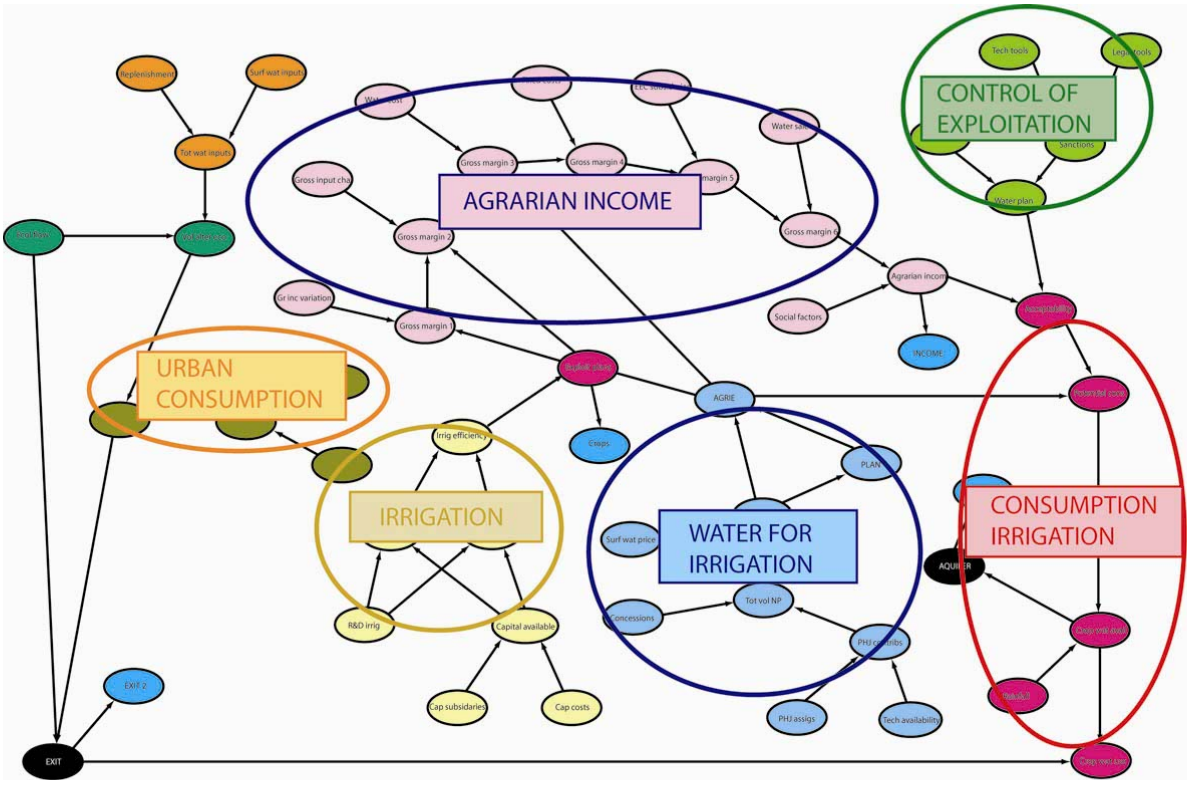
During construction of the network it is worth bearing in mind a few principles to make it easier to design and operate.

- *Grouping the variables:* Where variables can be represented as a group, then do so. Displaying the network as a series of colour-coded groups makes it easier to comprehend, something that is particularly important when trying to communicate information to stakeholders who may not be familiar with the environmental system. An example from Spain in given in Box 4.7.

- *Divorcing (separating child from multiple parents)*: Divorcing is a useful technique to simplify a model by minimising the number of links made to a single variable through the introduction of a mediating variable or variables.

  Consider the example taken from the Spanish network shown in Figure 4.4. The four links to the variable 'Exploitation plans' in (a), are reduced to two in (b), through the addition of two mediating variables (C_1 and C_2). The effect of these mediating variables is to 'divorce' the impacts of the original 4 interventions from the child node B. The mediating nodes combine variables that relate to the same theme. Thus A_1 ('Available surface water') and A_2 ('Groundwater abstraction') can be expressed as C_1 ('Maximum water available for irrigation').

Box 4.7 Grouping of variables in the Spanish Network



Placing the variables into colour - coded groups makes it easier to follow the logic of a network. The six main aspects of irrigation water use in the La Mancha aquifier in Spain are clearly identified by the process of grouping; without this the network would be much more difficult to comprehend.

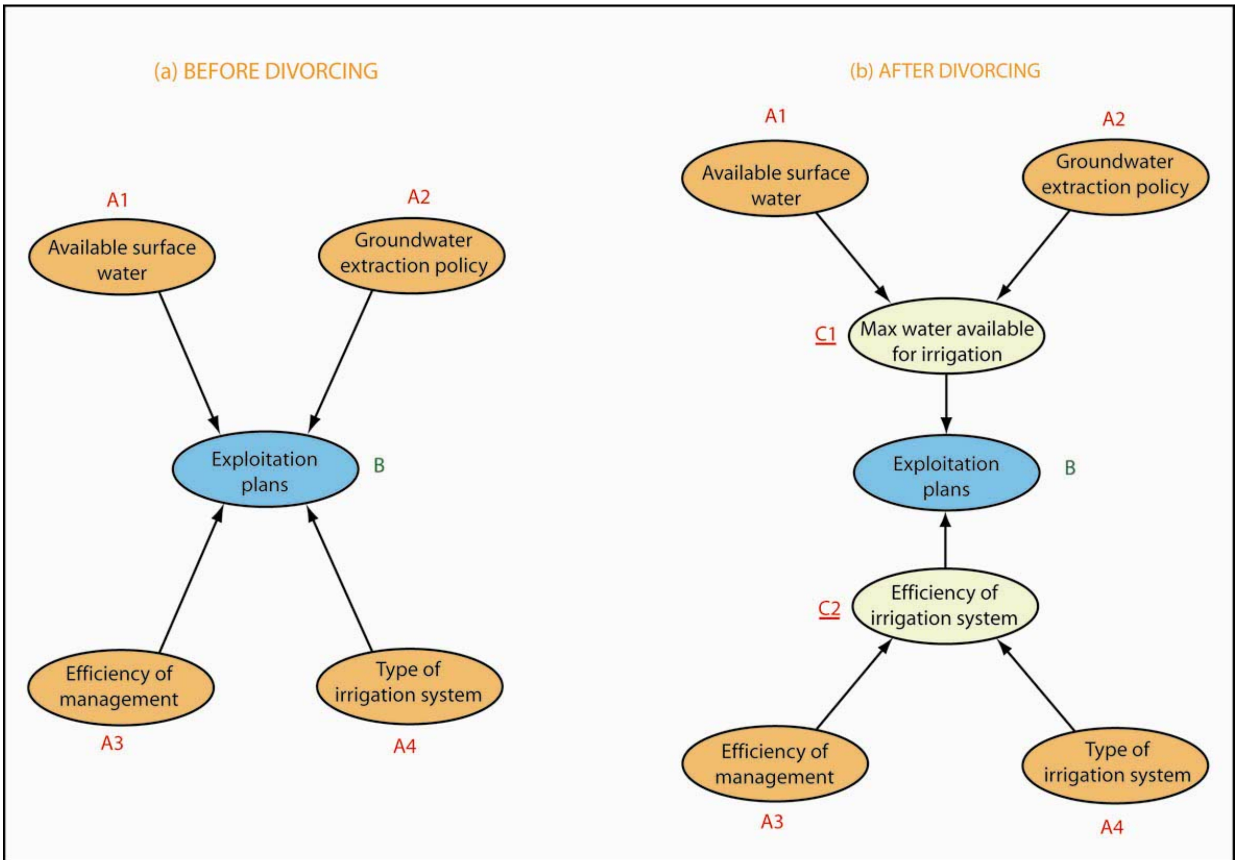


Figure 4.4 ‘Divorcing’, to reduce the number of links to a single variable

Likewise A_3 ('Efficiency of management') and A_4 ('Type of irrigation') can both be represented as 'Efficiency of irrigation system' (C_2).

With this modified design variable B only has 2 links, which makes the CPT much smaller and easier to construct. At first glance the second network looks more complicated with a larger number of nodes and links included, but the expanded configuration offers two great advantages: (a) it enables you to separate the impacts of different interventions and, (b) it makes the construction of the CPT tables much more straightforward.

- *The adder (reducing number of links to a child):* This is a technique designed to reduce the number of links to a child so that the effects of different parents can be assessed independently. It is similar to divorcing but does not rely on the creation of new mediating variables. Consider the example from part of the UK network shown in Figure 4.5.

  In the first case (a) the state of the variable 'Metered Consumption' (B) depends upon 4 other variables, A_1, A_2, A_3 and A_4. Variable B thus has 4 parents. This would be a logical way to express the situation, since the amount of water consumed domestically will depend on all 4 of the factors listed. However, we may be interested in the individual, rather than the combined, effect of the 4 factors on metered consumption. To accomplish this, additional nodes need to be created as shown in Figure 4.5(b). The effect of adding the extra nodes 'Metered Consumption (1), (2), (3) and (4), is to enable the effect of pricing, awareness campaigns and leakage reduction to be evaluated independently, for a specified number of houses.

  The effect is to separate out the impacts of the interventions from each other by *adding* one factor after another. Like 'divorcing' the 'adder' also has the effect of reducing the complexity of the CPTs.

3.3 Check network for consistency, logic and focus

Before moving on to the next step you should check to ensure the network is consistent, logical and does what you want it to do. It is much easier to change things at this stage before any of the CPTs have been entered. Some simple checks include:

1 Make sure your indicator variables properly represent the factors you wish to influence and are capable of being adequately quantified or measured. In the Danish flooding network, one of the main objectives was to monitor the ecological health of wetlands in the area. But instead of trying to quantify this complex concept, a variable showing water levels in the wetland was used; water level data is easy to collect and can be used as a surrogate to indicate wetland health. It is probably better to use an indicator variable with abundant data than one which may be a more precise description of your objective, but has less data and is more difficult to quantify.

2 Check all nodes without parents; these should represent either:
   - Actions
   - Scenarios
   - Implementation factors, such as funding
   - Controlling factors, such as rainfall

If they don't represent any of these then they probably require a parent, or perhaps are not necessary at all.

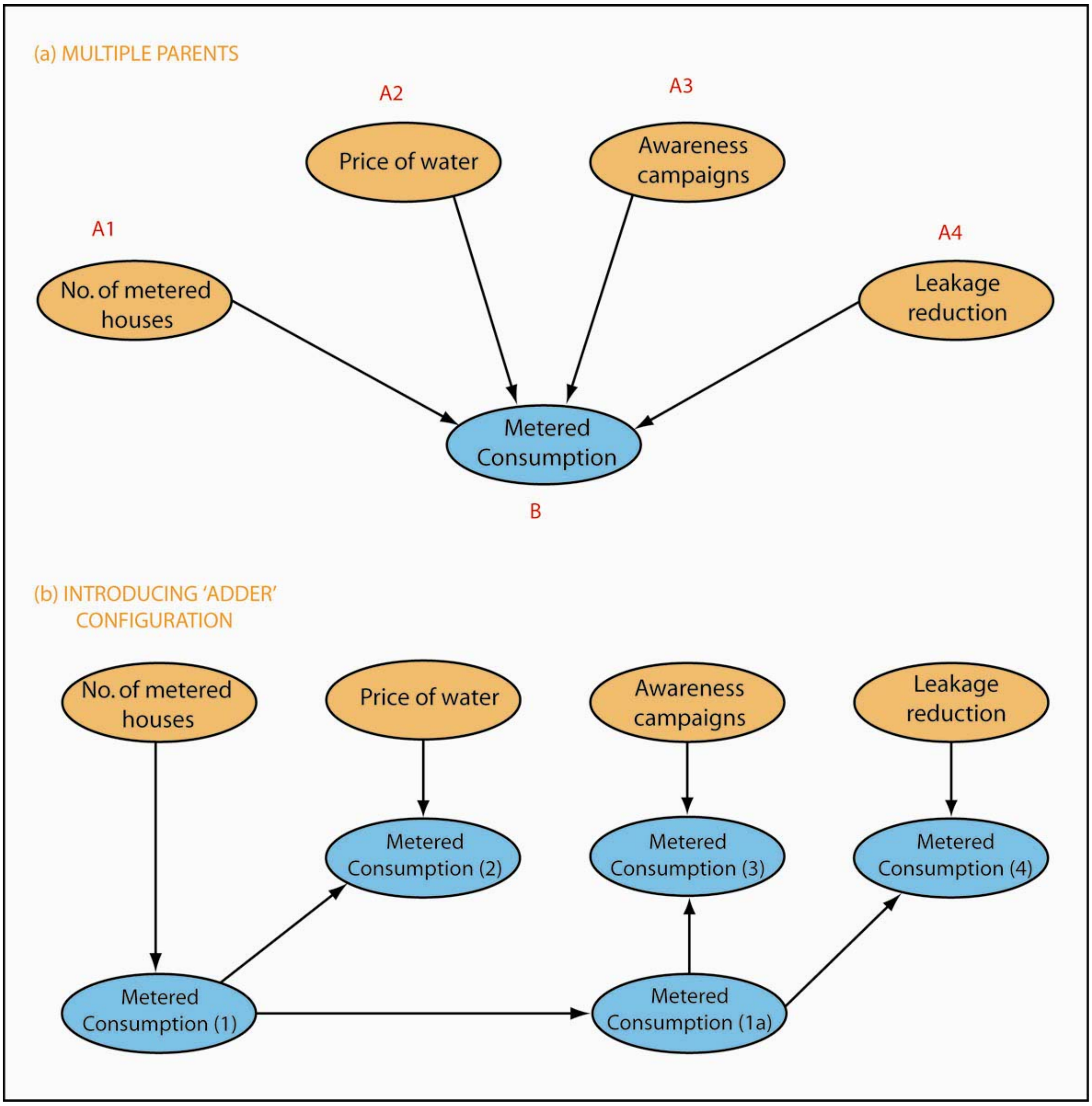


Figure 4.5 Introducing the 'adder' configuration to reduce the number of parents to a child and separate the effects of parent variables

Where a child node has more than one parent, it is possible that the impact of some parents on the child is influenced by the states of other parents. Nodes exerting this type of influence are defined as *modifying parents*. For example a child node 'Runoff' may have two parents, 'Rainfall' and 'Catchment Geology'. If the geology is clay, then the impact of rainfall on runoff will be high, because most of the rainfall will contribute to runoff as a result of the low permeability of the clay. However, if the geology is changed to sand, then the run-off contributed by the same amount of rainfall will be less, because of the increased permeability of the sand. In this case 'Catchment Geology' is said to be a modifying parent of 'Runoff'. Non modifying parents do not influence the impact of others on the child variable. Thus 'Crop yield' and 'Forest products' might be parents of 'Farmers' income', but in this instance changing the state of one parent will have no effect on the way the other parent impacts on the child; a poor crop yield does not affect the farmers' income from forest products and vice versa.

3 Check all nodes without children; these should represent either:

- Management objectives (indicators)
- Additional impacts as a result of selected implementations

If additional impacts (see Table 4.2) have been included, check to establish whether their inclusion is really necessary i.e. is changing the state of this variable going to influence your decision? If not, then it should be removed.

4 Check the d-separation properties (Appendix I). The d-separation properties of variables determine the way in which information is transmitted through the model, which clearly needs to be properly understood. A description of the principles of transmission of evidence using a number of variable configurations is given in Appendix I and need not be repeated here, but it is worth giving one example from the Spanish network.

In Figure 4.6 the variable C is dependent on variables A and B. This is a convergent connection. Using the rules outlined in Appendix I, it follows than if we have *any* evidence about the state of variable C (the total amount of water entering the system), then any knowledge of variable A (recharge) is bound to change belief about the state of B (surface water input), or vice versa; in other words A and B will be d-connected. In this case the principle is very obvious because A+B=C, so it follows that if we have some knowledge of C, and some evidence about A, then our belief about the state of B is bound to be influenced (Figure 4.6a).

By the same token, if there is no evidence at all about the total water input, then our belief about the state of B cannot be influenced by a knowledge of A and vice versa; in this case they are d-separated (Figure 4.6b).

If, when you start to run your network, variables change state when logically you might expect them to remain fixed, or do not change state when they should, this may indicate there is a d-separation / d-connection problem caused either by an unsuitable configuration or the inappropriate input of evidence. In this case a careful examination of the logic of the network is advisable.

5 Finally, does the network answer the questions being posed and will it provide you with the information to make an informed decision? If not, you need to consider whether you have set the right objectives and need to consider revising the network design and content.

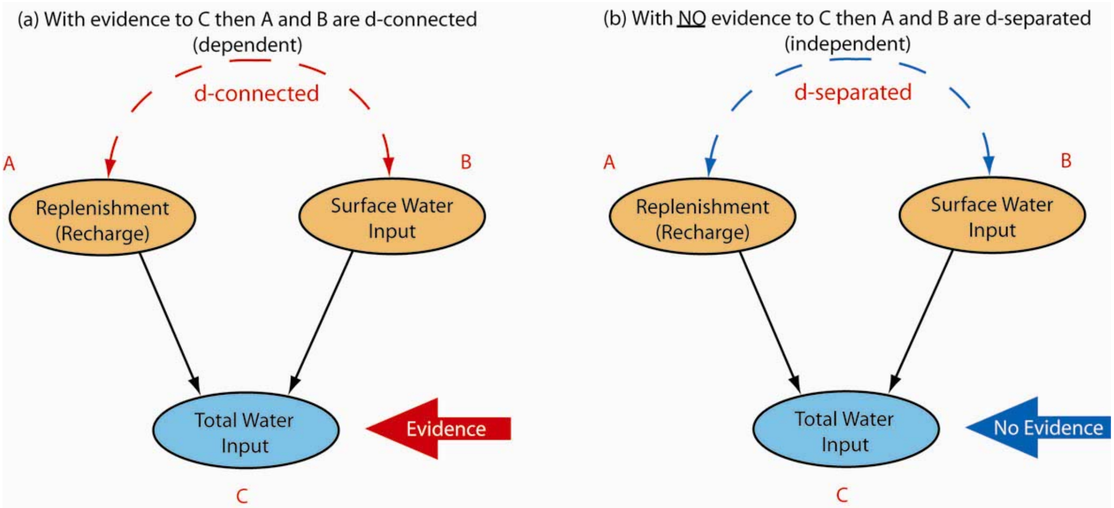


Figure 4.6 Checking for d-separation between variables; example variables taken from the Spanish network

Stage 3: Stakeholder involvement

This stage of the process gives the opportunity for stakeholders to influence the design of the network from the outset. The more the stakeholder can be made to feel a sense of ownership of the network the more likely you are to achieve a successful outcome, so make sure time and effort is spent on this stage.

3.1 Stakeholder input / feedback into the initial network

Having obtained a list of concerns, objectives and potential actions from your stakeholder group it is now time involve them in the construction of the first draft of the network. There are several ways to proceed:

1. You can construct the network yourself, based on information received from stakeholders and on your own expertise and knowledge, and then present it to the stakeholders for discussion
2. The network can be constructed together with the stakeholders at a meeting or series of meetings
3. An already completed network can be presented to the stakeholder group, variable by variable and link by link, for discussion and comment.

The first approach is the least time consuming, but is also the least transparent; stakeholders may feel they have been presented with a *fait accompli*. The second option is more transparent and gives stakeholders the opportunity to directly contribute to network design, though on the other hand it is likely to be a slow and perhaps contentious procedure. The third alternative offers a compromise between the first two. By constructing a network in advance, a great deal of time and effort can be saved, and serves the additional purpose of allowing the network constructor to become familiar with some of the pitfalls that may lie ahead with the construction. But then, rather than presenting the network to stakeholders in its entirety, it can be revealed variable by variable and link by link, giving the opportunity for discussion, comment and criticism. In this way stakeholders will gain a fuller understanding of the network, derive a sense of ownership, and have a real opportunity to influence the design. Which of these options is selected will depend on the time and resources available and the extent to which stakeholder involvement and transparency is seen to be paramount.

By and large it is preferable to adopt the second procedure whenever possible, though in many cases this may not be practical. It is likely that when you first begin to use networks you may not have the confidence to design them in the presence of stakeholders and prefer to use the first approach. This was the case for all the MERIT partners; however with experience the prospect of a more interactive approach becomes less daunting.

Even at this stage, before any data are entered into the networks, they can provide an excellent focus for the discussion of an environmental problem or strategy. At the end of a meeting, a clear graphical representation of the issues in the form of a network provides a much better synthesis of discussions than a list of points, or a few pages of notes.

3.2 Demonstrate completed network; further explanation

In Stage 2 you will have described the main features of Bayesian networks to the stakeholder group without going into too much detail. This may have been accompanied by a few simple examples. But now, before any data collection or input begins, is the time to explain in more detail how networks are constructed and the way in which they can be used to aid decision making. The best way to do this is to use some existing networks as examples; those included on the CD with these guidelines could be used. It is probably better to use real examples of environmental problems rather than the synthetic cases used by the Bn software packages.

The objective of this exercise is to not only demonstrate the way in which the networks function, but also how they can be used in the final decision making process. One of the comments from the MERIT case studies was that stakeholders did not realise the importance of the networks to the consultation procedure until it was too late for them to change anything.

At this stage it is also important to stress the limitations of Bns as well as their strengths, and that just like any other type of model they are only as good as the data fed into them.

3.3 Arrange system of dissemination to stakeholders / public

Good communication and dissemination of information is the key to the success of any decision making exercise involving stakeholders and / or the general public. Careful thought, therefore, needs to be given about the way in which this transfer of information is to take place.

If you decide to divide your stakeholders into two or more groups, there is first the problem of ensuring that information in the form of written minutes passes between the groups. In the Danish case the communication between the ‘professional’ and ‘citizens’ groups was seen to be inadequate and drew criticism, particularly from the ‘citizens’ group who felt that they were sidelined and not kept sufficiently well informed. One way to ensure better communication is have the same facilitator for both groups, or to have the decision maker attending both sets of meetings. In this way information can be passed from one group to the other verbally.

Even where only one stakeholder group has been convened there is still the issue of keeping the general public informed, where this is considered to be necessary. If sufficient funds and time exist, dissemination can be via regular newsletters distributed to local households, as was the case for the Danish network. Otherwise information can be passed through the local media (newspapers, radio), web sites, etc. Public information is particularly important in controversial cases where the impact will affect a large proportion of the population (e.g. decisions affecting domestic water supply). It is not so vital when the impact is restricted to a narrower interest group (e.g. decisions affecting the flooding of a small number of houses).

4.5 STAGE 4: Collect data from all available sources

Stage 4: Network construction

Once the design of the network has been agreed the next step is to collect the data for the probability tables that lie behind the variables. When collected, the data needs to be quality controlled and analysed for the requirements of the network. The three actions for stage 4 are:

- *Collect available data*
- *Quality control and analyse data in the light of network requirements*
- *Modify network if appropriate*

The main steps are shown in Figure 4.7.

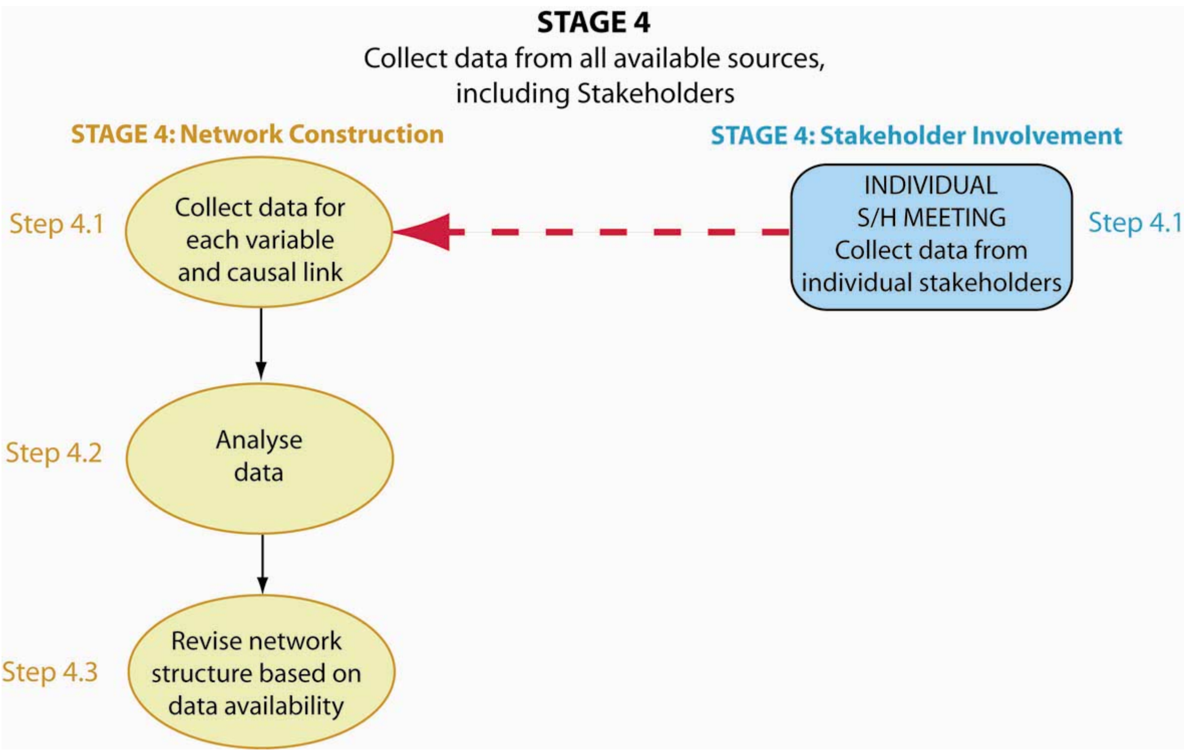


Figure 4.7 STAGE 4: Collection and analysis of data

4.1 Collect data

The main sources of data should have been identified during Stage 1; now is the time to collect the information. The data can come from many sources, but obviously the most reliable and complete data sets should always be sought. Bns can cope with different types of data, no matter how uncertain. There are 3 main sources of data:

1 *Measurements (e.g.* data from monitoring campaigns): If available this type of information is preferred, since provided it has been collected properly and objectively, there is no element of human judgement involved, removing at least one source of uncertainty. However, measurements are time consuming and expensive, and additional data are usually required from the other sources below.

2 *Model output:* Where monitored data is not available the output from models can be used as input to networks. Although not as reliable as measured data, it is possible to use models to generate large amounts of information to cover many cases, a feature that is useful for scenario setting. When using model output bear in mind that the results are only as good as the model and the data fed into the model; it is always advisable to run sensitivity analyses on model output to establish the scale of uncertainty attached to the result.

3 *Expert opinion:* Where little or no information exists, you may need to turn to 'expert' opinion. An expert is someone with specialist knowledge in a particular discipline (ecology, hydrology, economics etc), or of the local area, whose opinion can be trusted. In some cases it may simply be a stakeholder who, though not recognisably an expert in the conventional sense, has historical or recent knowledge that is of value.

Although extremely useful, the last approach should always be treated with some caution because stakeholders and experts, no matter how outwardly objective, will inevitably be biased. The bias may reflect a personal preference, because certain outcomes are more easily remembered, or for some other reason.

4.2 Quality control and analyse data

Data collected by any of the above methods will require manipulation before they can be entered into a CPT. Some of these techniques are discussed in Stage 6 (Construct probability tables) and in Appendices II and III.

Inevitably some data will have problems. These might include:

1 Data from stakeholders and / or experts may be inconsistent. Where this happens, an element of judgement is required by the network constructor. A number of options are available:

- More expert / stakeholder opinion can be sought until a consensus is reached
- A decision one way or another is made, based on the information available
- Where it is impossible to decide between two options both can be included in the network and the results of each used to help make a final decision

2 Data may be available for different periods of time for different data sets. This problem can be overcome to some extent if there is overlap of records allowing a relationship between the two sets of data to be established and the shorter record to be extrapolated. In the case where no overlap exists:

   - Use a relationship between the parameters that has been developed elsewhere in a similar environment and trust that it holds true for your region
   - Try to find a surrogate parameter. For example, in an irrigated region you may need data on borehole abstraction, but none exists. Instead it may be possible to estimate the area of irrigation using aerial or satellite photographs and from this calculate total abstraction using known irrigation rates per unit area
   - Look for an alternative network structure that does not require the use of the data set(s) in question

3 Sometimes data may be missing or incomplete. In this case the same options as for case 2 apply.

Every effort must be made to ensure that data is consistent, logical and as complete as possible. This may involve some manipulation or transformation of the data into a suitable format. For example in the UK network total domestic consumption in the catchment had to be scaled up from the amount of water typically consumed by one person per day in an average household. To scale up to the catchment level meant having to multiply the figures by the average number of people per house, and then by the total number of houses in the catchment.

Finally, it is particularly important to remember that data needs to be consistent in terms of the geographical area to which it applies. If your network is based on a catchment and all hydrological data applies to this area, it is important for all social, economic and other data to refer to the same area. Where such data is not directly available it is valid to use information from outside provided you can be sure that it is appropriate for the network region. Economic factors, for example, will be controlled more by political actions and market forces, rather than catchment boundaries, so that it may be necessary and valid to use information from further afield. It is less likely to be the case with many physical data, which tend to be more site-specific.

4.3 Review network structure

The search for data will inevitably reveal gaps, which if sufficiently serious, may require a revision of the network. In certain circumstances this will mean having to omit some variables. Where data is very scarce consideration should be given to not proceed with the network, although such a problem should have been identified in Stage 1.

In other cases data deficiency can be overcome simply by re-designing the network in a way that makes it possible to use the information that is available. In this way networks encourage the designer to think about problems in a more creative and imaginative manner than is normally required.

Stage 4: Stakeholder involvement

At this stage any data held by the stakeholders needs to be collected.

4.1 Collection of data from stakeholders

Although some of the information required for the CPTs is probably available as published reports, data sets or papers, it is likely that much will need to be provided by stakeholders, or more precisely the organisations the stakeholders represent. These organisations should be prepared to release information, except where it is commercially confidential. Trust will need to be built up between the decision maker and the stakeholder group over the sensitive issue of data accessibility.

Where only poor quality data exists or none at all, you may be forced to rely on expert opinion and knowledge. In the Danish 'compensation' case an agricultural economist was employed to provide the information for many of the variables in the compensation node, based on his wide experience of the region in question. For the UK 'demand' network a GIS expert was used to provide estimates of house numbers using post code areas, in the absence of publicly available housing figures. And in the Spanish case, input to a number of the agricultural variables in the network came from local farmers and landowners.

Sometimes individuals may hold privately owned data sets. In the Danish flooding study good use was made of a 9-year daily meteorological record collected by one of the stakeholders, but which had not previously been made public. Not all stakeholders will have this type of data, but some will have historical knowledge of the local area. Much may be from memory and difficult to verify, but this type of expert opinion can prove useful in the absence of recorded data. Also, secondary sources such as records of local events in newspapers can be useful; this might include reports of floods, water shortages, pollution incidents, record angling catches etc.

4.6 STAGE 5: Define the states for all variables

Stage 5: Network construction

With the network design complete and data assembled, the next step is to assign states to each variable (Figure 4.8). This is an important task that needs careful consideration. The evaluation of whether a particular intervention will be successful or not depends upon the way that probabilities are distributed through the states that

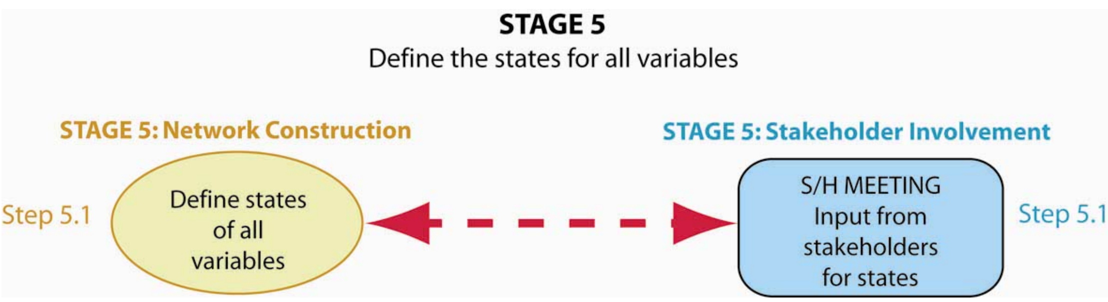


Figure 4.8 STAGE 5: Define states for all variables

have been set. Badly formulated states can result in a misleading impression of a particular outcome.

5.1 Define states for variables

Types of state
When assigning states you need to consider not only the current condition of the variable, but how that may be changed under different management scenarios or actions.

States only need to be defined where the node representing the variable is ‘discrete’. Where the node is ‘continuous’ the state is described as a ‘Gaussian (normal) conditional distribution function’, in which the user defines the mean and variance rather than a set of specific states. In the vast majority of cases, however, discrete nodes will be used.

Most commercial Bn packages allow you to describe the states of variables in one of 4 ways; as labels, numbers, intervals, or in a Boolean format. Some examples, taken from the MERIT networks, are shown in Table 4.4.

Type of State	**1. Labels**	**2. Numbers**	**3. Intervals**	**4. Boolean**
Node name	*Level of control*	*Compensation (Dkr ha^{-1})*	*Max water for irrigation (hm^3)*	*Water Saving campaign*
	High	0	0-20	Yes
	Medium	500	20-40	No
	Low	1000	40-60	
		1500	60-80	
		2000	80-100	
		2500	100-112.5	
		4400		

Table 4.4 Types of state; examples from MERIT

The use of *labels* is illustrated by the variable ‘Level of control’, taken from the Spanish Bn (Table 4.4). In this example three states are described as ‘high’, ‘medium’ and ‘low’. The variable describes the degree of control exercised by the authorities over the abstraction of water for irrigation. No values are attached to the states because it is difficult to quantify ‘degree of control’. But labels can also be used for states that can be quantified. A variable, ‘Income’, could have the states high, medium and low, but each of these might represent an actual figure or range of income. Labels are used for clarity. They are less confusing than numbers, and make it easier to take in information at a glance.

In some cases, it may be more appropriate to describe the states using *numbers*. The variable ‘Compensation (Dkr ha^{-1})’ describes the different levels of annual compensation potentially available to farmers in Denmark for reducing pesticide use per ha. Seven states, ranging from zero to 4400 Dkr, are assigned to the node. Note that the step between numbers need not be regular. The size of step is left entirely to the user.

It is also possible to use *intervals* rather than single values. This is illustrated by the ‘Max water for irrigation (hm^3)’ variable taken from the Spanish network. Here the volume of water is represented as a series of intervals covering the likely range to be encountered. As in the case for numbers, the intervals need not be regular or fixed.

A fourth type of state is represented by the ‘Water saving campaign’ node. This example, taken from the UK network, is used to illustrate the ***Boolean*** type of state. Here there is a choice between one of two conditions; in this case it is either ‘yes’ (there is a water saving campaign) or ‘no’ (there is not). This type of state is useful when representing a possible action, which may or may not be taken, or to identify whether something is ‘true’ or ‘false’.

Finally it should be noted that the labelled state can be used to represent any of the other 3. For instance the Danish networks use labels for all the states of all variables; this allows a mixture of intervals, numbers and letters to be used. In the flooding network the states are described in the following form 100-500 l/s (i.e. interval and text together). It is up to the user to decide the best option for their particular case.

Some rules for selection of states

1 *Minimise number:* The number of states should be kept to a minimum for two reasons: (a) for clarity, and (b) to simplify completion of CPT tables.

Sometimes, though, a large number of states cannot be avoided. A good example is provided by the Spanish network. The large window in Figure 4.9 shows total abstraction from the La Mancha aquifer in $hm^3\ yr^{-1}$, which ranges from 60 to 700 $hm^3\ yr^{-1}$. Rather than attempt to compress this wide range of flow into 3 or 4 states, the designer has instead opted to retain sensitivity and to represent flow as 23 intervals of 20 $hm^3\ yr^{-1}$, plus one of 200 $hm^3\ yr^{-1}$.

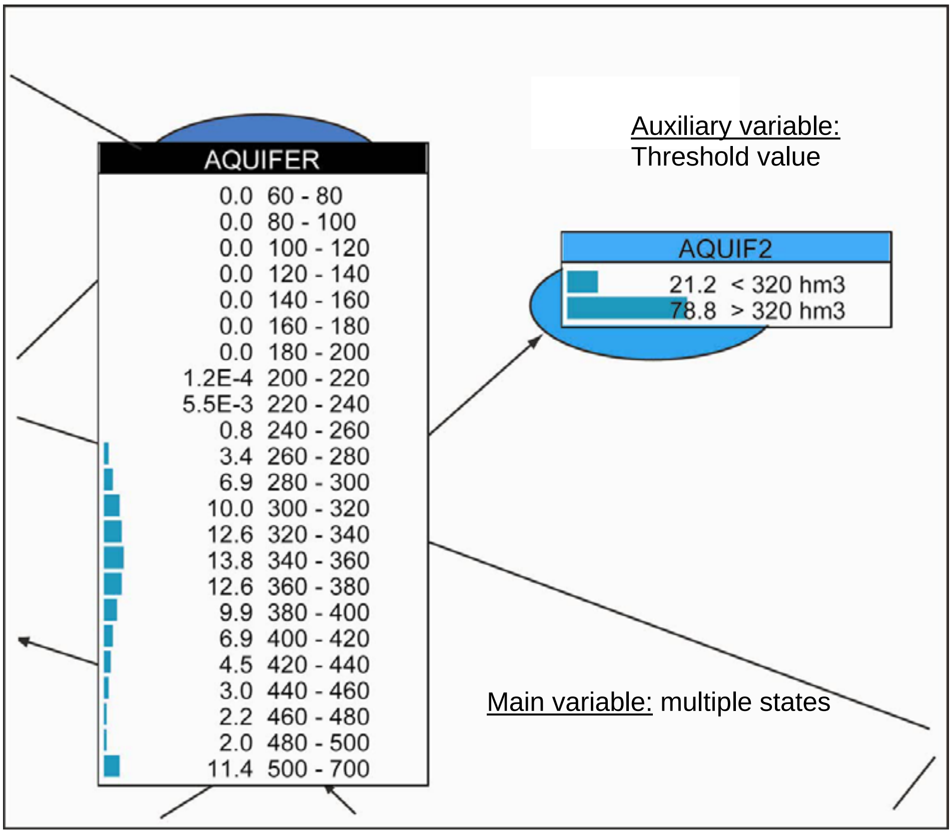


Figure 4.9 Variable with a large number of states linked to an auxiliary variable using a threshold value

2 *Auxiliary nodes*: Although the multiple states shown in the large window in Figure 4.9 serve to retain the sensitivity of the information, the probability distribution is rather wide and difficult to interpret without close scrutiny. Where this is the case it is useful to attach an auxiliary node, which shows the same information, but uses only two states. In the example the auxiliary node has one state above 320 $hm^3\ yr^{-1}$, and one below. This abstraction rate has been selected as a threshold above which sustainable aquifer development is not possible. From the parent variable it is difficult to estimate whether this threshold value is being attained or not. But a quick glance at the auxiliary (child) variable confirms that for this particular scenario there is a 79% probability that abstraction exceeds the threshold, and that as a result the situation is unsustainable.

3 *Relevance of states:* It is important to bear in mind that states play a crucial role in communicating your ideas. Try to create states that are intelligible and relevant to the stakeholder and not simply an abstract academic or technical concept.

Two examples of states intended for different types of stakeholder are shown in Figure 4.10. The first case (a) is taken from the Italian network constructed by CESI to evaluate the environmental performance of different water management policies on aquatic life in the Vomano catchment. This is a specialist network intended for stakeholders with an interest and knowledge of aquatic ecology; the list of fish types (fario, lasca) forming the different states would, therefore, be perfectly intelligible to this audience. The second case (b) is from the Danish compensation network, where many of the stakeholders are farmers. For this group the most appropriate way to express results is in monetary terms, something they understand very well. The network designer, therefore, has chosen to express the suitability of land for hunting and fishing as a monetary value (Dkr) per hectare.

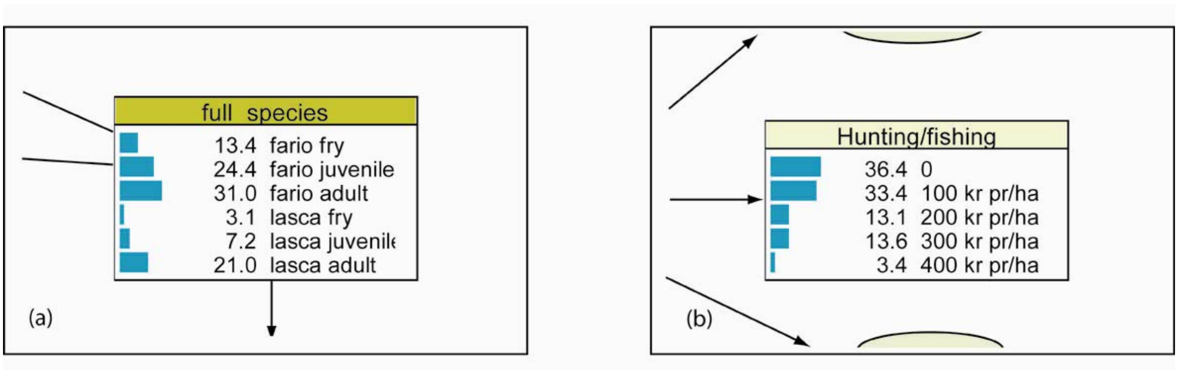


Figure 4.10 Different types of states intended for different types of stakeholder

4 *Exhaustiveness and exclusiveness:* The states assigned to a variable must be exhaustive and exclusive. By exhaustive we mean that the selected states must cover all the conditions the variable is likely to encounter. In Figure 4.9 if the discharge from the aquifer is capable of exceeding 700 $hm^3\ yr^{-1}$, then the states shown are not exhaustive, and more need to be added to cover all possible future events.

Exclusiveness means that if the state of a parent variable is known it must be exclusive of the other states. For instance, a variable called ‘Actions’, might be assigned 2 states, 'Build a dam' and 'Construct boreholes'. However, these states are not exclusive because it is possible to do both simultaneously. Under these circumstances it is necessary to create two variables 'Build a dam' and 'Construct boreholes' and to have Boolean states for each (i.e. yes and no).

5 *Creation of separate variables for different states:* Be careful not to create separate variables for different states of the same variable. This will help keep down the size of the network. As an example in the Danish compensation network the variable 'Animal/vegetable', which refers to whether the farm has animals or crops, could have been expressed as two separate nodes, but such an action would have been unnecessary.

Stage 5: Stakeholder involvement

The selection of states needs to be done in collaboration with stakeholders since the success or otherwise of a particular decision is made on the basis of the probability distribution over the states selected. Inappropriate selection of states will make it more difficult to gain acceptance for decisions.

5.1 Selection of states for variables

Selection of states can best be done at a group meeting. The alternative is to visit stakeholders individually, but this is time-consuming and does not allow for the exchange of ideas that takes place in a group.

Stakeholders should be asked whether they prefer the states to be expressed quantitatively (a number or range) or qualitatively as a description. In the Spanish network 'Water Cost' has been given 3 states '0.05-0.07', '0.07-0.1' and '0.1-0.12', the units being Euros per m^3. However, these ranges can be confusing to follow, and some stakeholders might have preferred to express the states as 'low', 'medium' and 'high', where each description refers to a specific price range (Figure 4.11).

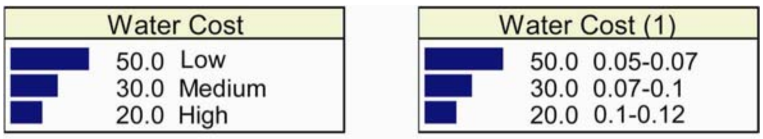


Figure 4.11 When states are labeled they are easier to follow

For some variables the most obvious set of states is not necessarily the most easy to follow. When this happens it is worth asking the stakeholders if the outcome they are looking for can be expressed in a more clear way. For instance in Figure 4.12, taken from the UK network, the same outcome is expressed in two different ways. The selection of variable and the type of state should be discussed and agreed with the stakeholder group.

Remember, allowing stakeholders to select their preferred means of expressing the outcome, will make it easier for them to understand and accept the results.

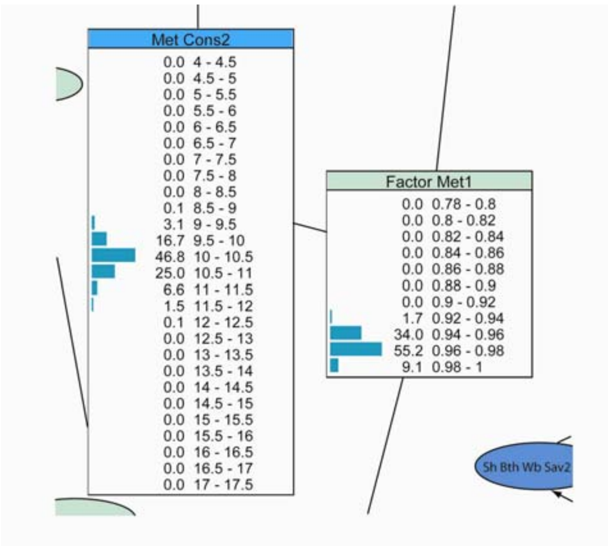


Figure 4.12 Different ways of expressing an outcome. The window on the left shows total household water consumption in Mm^3, following a price change. The window on the right shows the same effect, but expressed as a fraction of water use before the price change; it could also have been expressed as a percentage of water saved

4.7 STAGE 6: Construct conditional probability tables

Stage 6: Network construction

The next step is to complete the conditional probability tables that lie behind each of the variables. This is a task that requires care and patience because the CPTs are the core of the network; if the CPTs are wrong, the network output will be equally wrong. As well as being one of the most important, it is also the step that causes most difficulty for the network designer and for the stakeholders who provide information for the table.

For each variable, the complexity and size of the CPT depends on the number of parents and states the variable has; the more parents and states, the greater the size and complexity of the tables. This is a major reason to try and limit the number of parents and the number of states as described in Stages 3 to 5. The procedure used to complete CPTs depends on the type of data being used. There are two main types you are likely to have:

1. *Stakeholder opinion / knowledge and expert opinion:* Sometimes the data linking two variables may be scarce, or none existent, particularly when the link is difficult to quantify. In these cases you will need to turn to local stakeholder knowledge, or expert opinion. Types of variable that are difficult to quantify include those dealing with social or cultural issues, but economic factors can also be problematical. As an example, you might be concerned about the way in which the value of riverside property is influenced by the condition of the river. Poor quality water or a lack of flow for instance would cause prices to fall, but by how much? This type of information is most likely to be obtained not from data sets but from people with knowledge of the housing market in the area, such as local estate agents. Where this is not available, then 'expert opinion' from a more distant estate agent, or a housing expert, might be used.

2. *Monitored data and model data sets:* These data sets may be very large, particularly those generated by models. Good examples of variables for which large *data sets* are frequently available are daily rainfall and river flow. On the other hand data for a variable such as groundwater level often have to be simulated using a

model. Where a model is used, as much data as necessary may be generated to cover as wide a range of conditions as possible.

If you have the first type of data and are relying on stakeholder input or expert opinion, the data must be carefully elicited and input into the tables *manually*. Manual input may also be appropriate where there is a small amount of monitored or modeled data that can be handled easily. Some simple software has been developed by Cain (2001) to help in cases of manual input where tables are large. A copy of this software is included on the CD that comes with these guidelines. An example of manual input to CPTs is given in Appendix III.

If on the other hand you have the second type of data, with large data sets to handle, the only practical way to enter the probabilities into a CPT is to use a facility available in most Bn software known as *parameter learning;* this technique interrogates the available data set to specify probabilities. An example of this type of procedure is given in Appendix II.

The steps for stage 6 are shown in Figure 4.13.

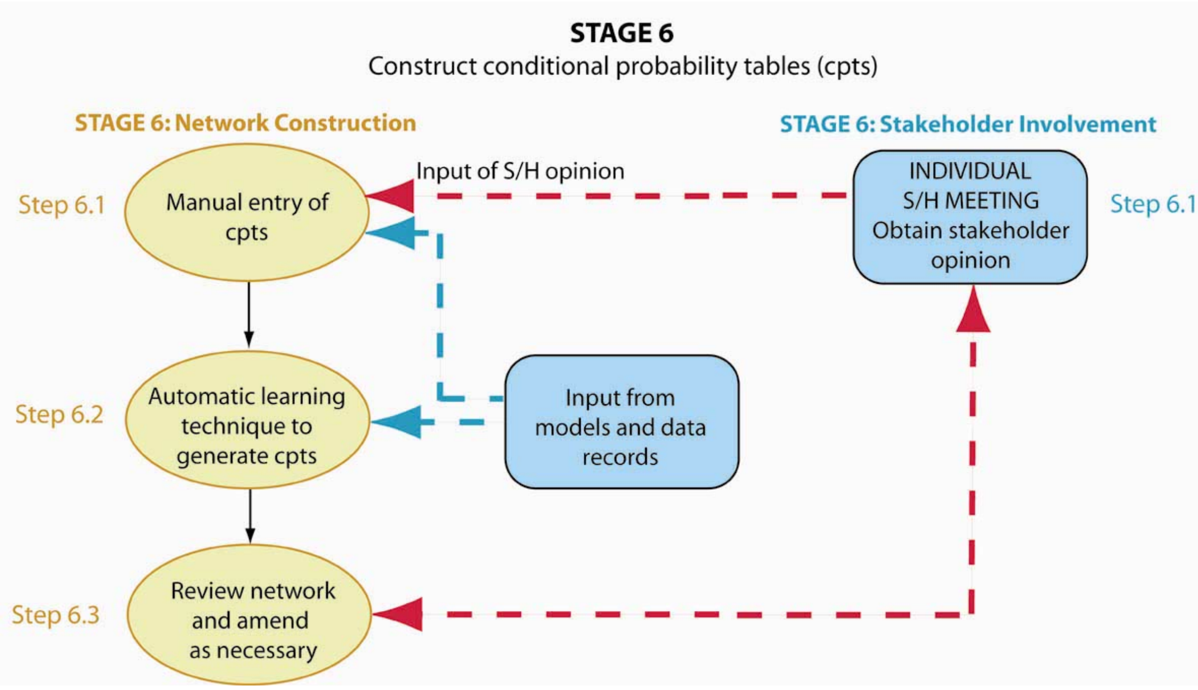


Figure 4.13 STAGE 6: Construct conditional probability tables (CPTs)

6.1 Manual entry of CPTs using expert opinion and/or small amounts of data

Manual entry of data into CPTs is necessary either when stakeholder or expert opinion is used or where the amount of data is small and capable of being handled manually.

For many links, there may not be any data or model output connecting variables, particularly where you are dealing with factors that are difficult to quantify such as social issues, the 'value' of environment, or perceived risks or benefits. Under these circumstances the probabilities for the CPTs need to be provided either by stakeholders with local knowledge, or by experts who may not be local but are specialised in a particular field. The drawback of using stakeholder input or expert opinion is the risk of bias or selective memory. But there are also advantages. Stakeholders may provide historical information that would otherwise be lost, or provide observations or insights not apparent from the data available. There is also the advantage that involving stakeholders increases the sense of ownership of the network.

When data is input using contributions from stakeholders or experts, there are a number of things to note:

(a) *Minimise the size of the table:* Information obtained from stakeholders or experts has to be entered into the CPT manually. To complete the table you need to ask the stakeholder or expert the probability of the child being in each state given all the combinations of the parent states, so the smaller the table the better. This can be achieved by keeping the number of parents and states as small as possible (see Stage 3).

(b) *Expert input:* Using an expert to input probabilities means that values can be entered directly into the table since he or she is likely to be familiar and comfortable with the concept of probability. It also means that more complex tables might be considered when experts are being employed. A good example is offered by the Danish compensation network. The variable 'Farm Economics' has 3 parent nodes: 'Farming Contracts', 'Animal / Vegetable' and 'Land Use'. Between them the parents and child have a combined total of 21 states (8+8+3+2) which meant that the CPT for 'Farm Economics' required 384 values. However, despite the size and complexity, the CPT was successfully completed in conjunction with an expert familiar with the economics of agriculture in the area.

(c) *Stakeholder input:* Some stakeholders may be experts in their own right. In the UK example, one of the stakeholder organisations was a private water company, but also an expert in the field. In these cases direct input of data into CPTs is quite possible as described in point (b). However, other stakeholders, although possessing considerable local knowledge, may not be familiar with expressing data in terms of probability making it necessary to devise carefully formulated questions in order to elicit the required values. Thus instead of asking a farmer the probability of his income being low, given a poor crop yield but a good return from forest products, you could ask how many years out of 10 his income would be low if crop yield was poor but the return from forest products good. Alternatively, rather than asking for probabilities, you may get a better response if you replace probabilities by descriptions. So instead of asking the farmer for a probability you could ask whether an event was 'certain', 'likely', 'not likely', 'never' and so on; you can then convert these to probabilities at a later stage. Putting the question in a context to which the stakeholder can relate, or using terms with which they are more familiar, may make it easier to obtain the necessary input for the CPTs.

(d) *Quantification of variables:* It is often difficult to quantify abstract variables such as the amenity value of landscape, although these types of variables often dictate whether or not a particular management action is acceptable. What social value does a landscape have? And what constitutes acceptable? To environmentally conscious members of the public a damaged wetland is unacceptable, whereas to others it may not be a cause for concern. Another option is to attempt to place a monetary value on the amenity. This is the approach adopted in the Danish compensation network where instead of trying to frame the ‘recreational’ value of land in terms of social acceptability, it has been expressed as a value in Dkr per hectare.

In cases where expressing variables in financial terms is not possible then again information for the CPTs needs to be elicited from stakeholders through carefully framed questions. Suppose a management action has an impact on river flow and you have to assess whether this will affect the social and amenity value of the river. The presence of healthy river provides a difficult-to-quantify social value, yet one that is crucially important. How do you try to quantify this for a river in terms of its acceptability as an amenity in order to fill in the CPT? One approach is to identify a threshold below which a particular situation is not acceptable. In the case of a river the threshold could be framed in terms of the flow volume past a specified point, the level of the river, or alternatively the number of fish along a particular stretch. The acceptability of reduced flows in a river for instance could be assessed by asking such questions as if the river is dry for ‘x’ weeks a year is this acceptable? If the flow is sufficiently low to reduce the fish population by ‘x’ percent, is this acceptable? Stakeholders will find it straightforward to answer this type of question and the child variable will only have two states, ‘acceptable’ and ‘not acceptable’, which will restrict the size of the CPT making it more easy to complete.

More information about the procedure to elicit data from stakeholders is given in Cain (2001) a copy of which is given on the CD accompanying these guidelines. More information is also given in Appendix III.

6.2 Parameter learning to complete CPTs using large data sets or model output

The concept of experience
Where measured data sets or data from model output is available, the probabilities for CPTs need to be extracted from this information. Probabilities are derived using the concept of ‘experience’; the procedure can best be explained by using an example.

Figure 4.14 shows part of the Danish ‘compensation’ network relating to groundwater quality. Two parent variables, ‘Shallow gw quality’ and ‘Point sources’, link to a child, ‘Deep gw quality’. The CPT for the child (‘Deep gw quality’) is shown in the window at the top.

The set of states assigned to the ‘Point sources’ parent is given in the top *row*, to the right of where the variable name appears. It has 3 states ‘None’, ‘Moderate’ and ‘High Intensity’. In the row below, the set of states for the ‘Shallow gw quality’ parent appears. Again there are 3 states (all in µg l^{-1}), ‘<0.01; 0.01-.1 and >.1. These are repeated for each state of ‘Point source’ in the row above. This gives a total of 9 combinations for the two sets of states for each parent. Finally the set of states for the child variable (‘Deep gw quality’) is listed in the extreme left hand *column*. Once again there are 3 states, (all in µg l^{-1}), ‘<0.01; 0.01-.1 and >.1. With 9 possible combinations of parent states, and 3 for the child, this means the CPT for ‘Deep gw quality’ has a total of 27 combinations.

Each number in the table (white boxes) refers to the probability of the child (‘Deep gw quality’) being in a specified state for a given combination of parent states. The probabilities are calculated using data which links the quality of deep groundwater to the quality of infiltrating surface water and the number of point pollution sources in the area.

For example, if the state of ‘Point source‘ is Moderate and ‘Shallow gw quality’ is <0.01, the probability of the child being in the state 0.01-0.1 (middle figure, left hand column) is 0.6 (60%). This figure is calculated from the available data set by counting the number of times the child is in this state for the particular combination of parent states (i.e. Moderate and <0.01), then comparing it with the total number for that parent combination. Each piece of information that applies to one combination of the parent states is known as a ‘case’; the count of the number of cases for each combination is termed ‘experience’.

Let us again take the parent combination ‘Moderate’ and ‘<0.01’. The total number of cases reported for this combination might have been 20 of which 6 fell into the child state ‘<0.01’, 12 into ‘0.01-0.1’ and 2 into ‘>0.1’. Converting to probabilities this gives the 0.3, 0.6 and 0.1 distribution shown in the table.

Deep gw quality

Edit Functions View

Point sources	None			Moderate			High intensity		
Shallow gw...	< 0,01 u...	0,01 - 0,...	> 0,1 ug/l	< 0,01 u...	0,01 - 0,...	> 0,1 ug/l	< 0,01 u...	0,01 - 0,...	> 0,1 ug/l
< 0,01 ug/l	1	0.3	0	0.3	0.1	0	0	0	0
0,01-0,1 ug/l	0	0.7	0.3	0.6	0.7	0.7	0.8	0.7	0.5
> 0,1 ug/l	0	0	0.7	0.1	0.2	0.3	0.2	0.3	0.5

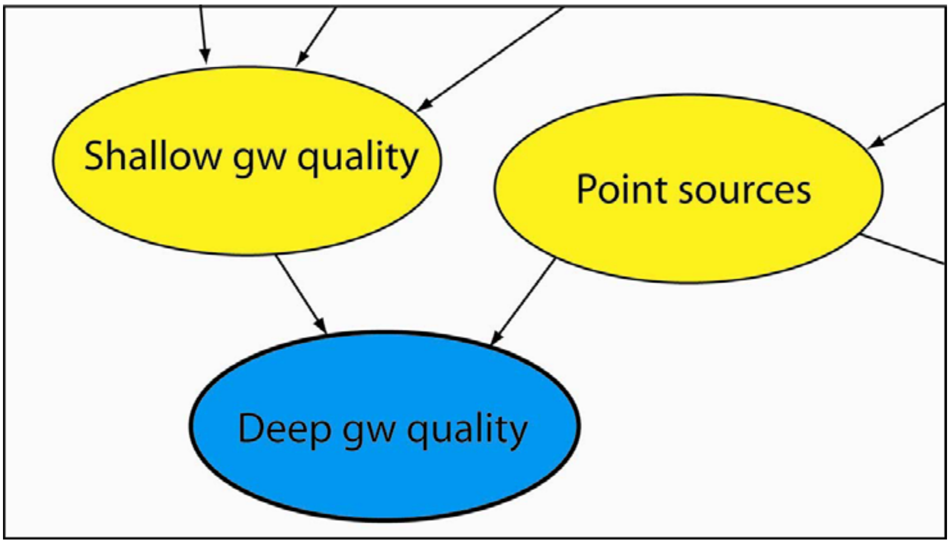


Figure 4.14   Typical CPT compiled using a data set. The states in the column on the left are those for the child variable states (Deep gw quality); the states shown in the two rows at the top are for the parent variables (Point sources and Shallow gw quality).

The more cases available, the more ‘experienced’ become the probabilities in the CPT. Where only one or two cases have been used to obtain a probability the addition of an extra case in the form of new information will have a large effect on the result. On the other hand where twenty or thirty cases have been used, any new information will have minimal impact. You will obviously have much more confidence in a situation where a large number of cases are used. As a rule of thumb it is a good idea to have at least 20 cases for each combination. If insufficient cases are available for certain combinations it is worth reducing the number of states in one or more variables to accommodate more cases per combination. Should the number of cases be so small that only a few cases are available for each combination, then probabilities should be entered manually in the manner described in the section dealing with the construction of CPTs using expert opinion.

Learning conditional probabilities from data or model output
Bayesian network learning is the automatic process of determining a suitable network given data presented in the form of cases. Traditionally the learning process is divided into two parts, *structure learning* and *parameter learning*.

Structure learning is where you allow the software to determine the location and direction of links between nodes, based on the observed dependence / independence of variables. When using the structural learning option the operator simply inputs the data containing all the different variables that are of interest, and the software then allocates the most appropriate links based on the data available. This is a useful facility when you are uncertain about the relationships between variables; it also has the advantage of removing operator bias. However, where the conceptualisation of environmental systems are concerned, it is likely that in most cases the cause and effect between variables will already be known, so this learning facility will not often be required, though of course it could be used to confirm your ideas about the way the system operates.

Parameter learning, on the other hand, will almost certainly be needed. Parameter learning calculates the conditional probabilities for variables, given the link structures and data. In other words it will automatically populate a CPT given the relevant data sets. Where this technique is used, it is necessary to have sufficiently extensive data sets to generate enough cases (experiences) for the resulting probability distributions to be robust. One of the most commonly used methods for parameter learning is the Expectation-Maximisation (EM) algorithm.

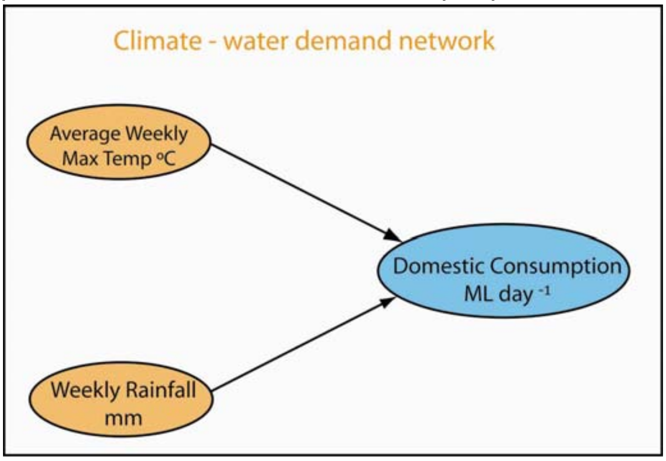
The EM algorithm
If you have a large data set it is likely that sufficient 'cases' can be generated to provide values for all parent combinations. In these circumstances the network is said to be 'complete', and the calculation of probabilities becomes a straightforward process as described above. However, in reality, data sets are often patchy and there is a good chance that values for all parent combinations cannot be assigned, leaving gaps in the CPT table. The network is then said to be 'incomplete'. When this occurs, explicit determination of conditional probabilities is not possible. Instead, an iterative method is required to complete the probability distributions. One of the most commonly used methods is the (EM) algorithm (McLachlan and Krishnan, 1996).

The EM algorithm is used to estimate CPT values iteratively from observed data, starting from some initial estimate of the values. Each iteration comprises two steps, an Expectation (E) and a Maximisation (M) step. The E step finds the expected distribution for the missing values based on the known (observed) values and initial estimates. The M step then re-calculates the values to find those with the maximum likelihood, given the distribution derived from the E step. Successive iterations improve the likelihood until changes become progressively smaller and smaller and fall below a convergence threshold.

It is not necessary to be familiar with details of this technique since most Bn software will provide a wizard menu to enable you to enter the data required. Should more information on EM learning be needed you can consult the help files of the Hugin or Netica software, or publications such as McLachlan and Krishnan (1996). The best way to illustrate EM learning is to describe an example using the Hugin software.

Example of EM learning

Figure 4.15 shows a small network designed to predict the weekly domestic water consumption in a catchment given the weekly average maximum temperature and weekly rainfall. The water company was interested to know whether it would be possible to forecast consumption based on predictions of weather conditions provided by the meteorological office. Intuitively there should be a relationship between temperature and consumption because in warmer weather people tend to wash, shower and drink more often. Similarly during periods of low rainfall more water will be used in the garden. The catchment is in the UK, so the rainfall and temperature data are for the summer months only; in winter, consumption is independent of temperature and rainfall. Other factors such as school holidays, pipe leakage, types of housing etc will influence consumption, but these are not taken into account in this network. As a result, there will inevitably be a degree of uncertainty in the result.



The local water company provided a data-set of weekly values for maximum temperature, rainfall and domestic consumption for a seven year period. Excluding the winter months, this leaves a total of 182 records, of which the first ten are shown in Table 4.5.

Figure 4.15 Climate – domestic water demand network

Max Temp ^{0}C	Domestic consumption ML day^{-1}	WeeklyRainMM
10.21	211.3	41.1
10.22	212.2	22.7
13.92	220.7	0.0
12.28	220.5	9.5
13.56	227.7	0.0
15.03	229.8	0.6
14.20	220.9	4.0
14.36	225.5	25.9
15.54	213.1	9.1
16.47	218.1	33.0

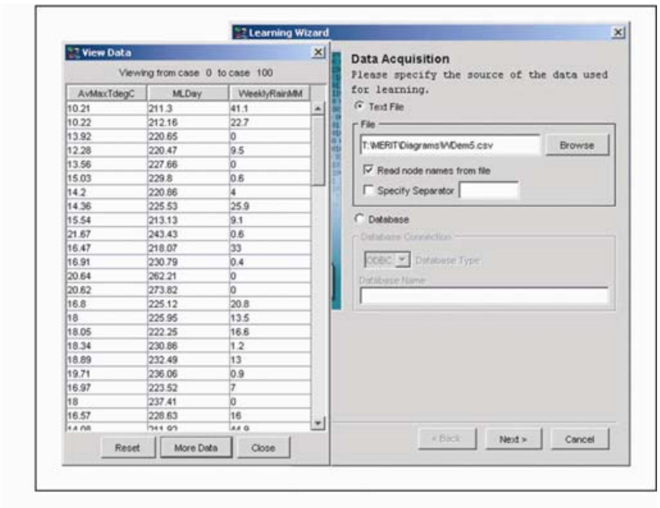
Table 4.5 First 10 lines of the climate-demand network; total number of records 182

Max Temp ^{0}C	Domestic consumption ML day^{-1}	WeeklyRainMM
0-15	210-220	0-5
15-20	220-225	5-10
20-25	225-230	10-15
	230-235	15-20
	235-240	20-30
	240-250	30-100
	250-300	

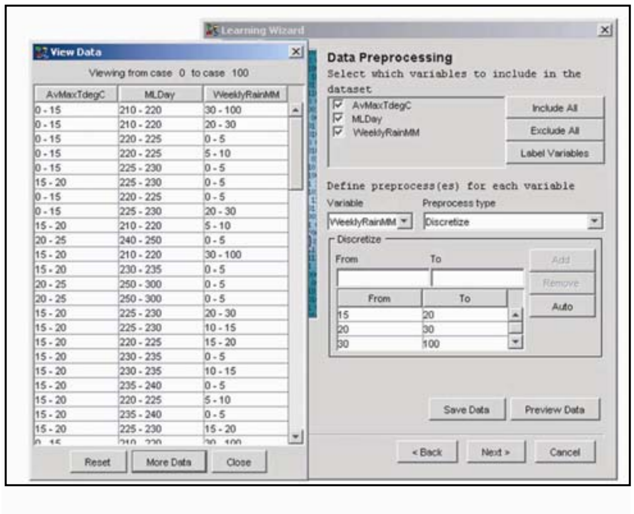
Table 4.6 The states for the 3 variables in the climate – demand network

The states for each variable are listed in Table 4.6. For the child node, ‘Domestic consumption’ seven states were selected to represent the range of anticipated demands; for the two parent nodes there were three states for ‘Max Temperature’, and 6 for ‘Weekly Rainfall’. The ranges of the child node were selected to ensure that at least some observed values fell into each range.

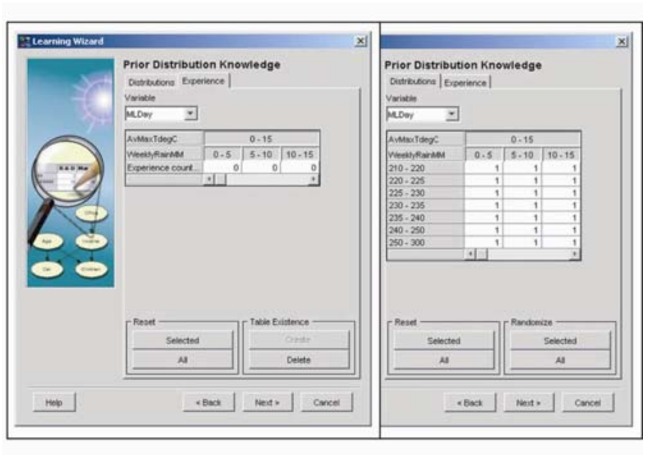
With this number of states the CPT for the child node requires 126 values to be estimated, making manual calculation and entry tedious and time consuming. Fortunately the Bn software allows the CPT to be completed automatically using the EM learning procedure. The procedure can best be illustrated through a series of screen dumps taken from the Hugin package using data for the climate – demand network (Figure 4.16).



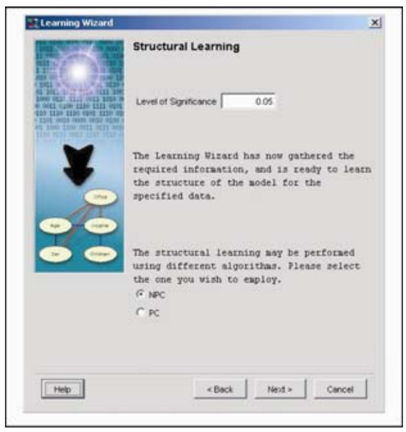
(a) Select data file



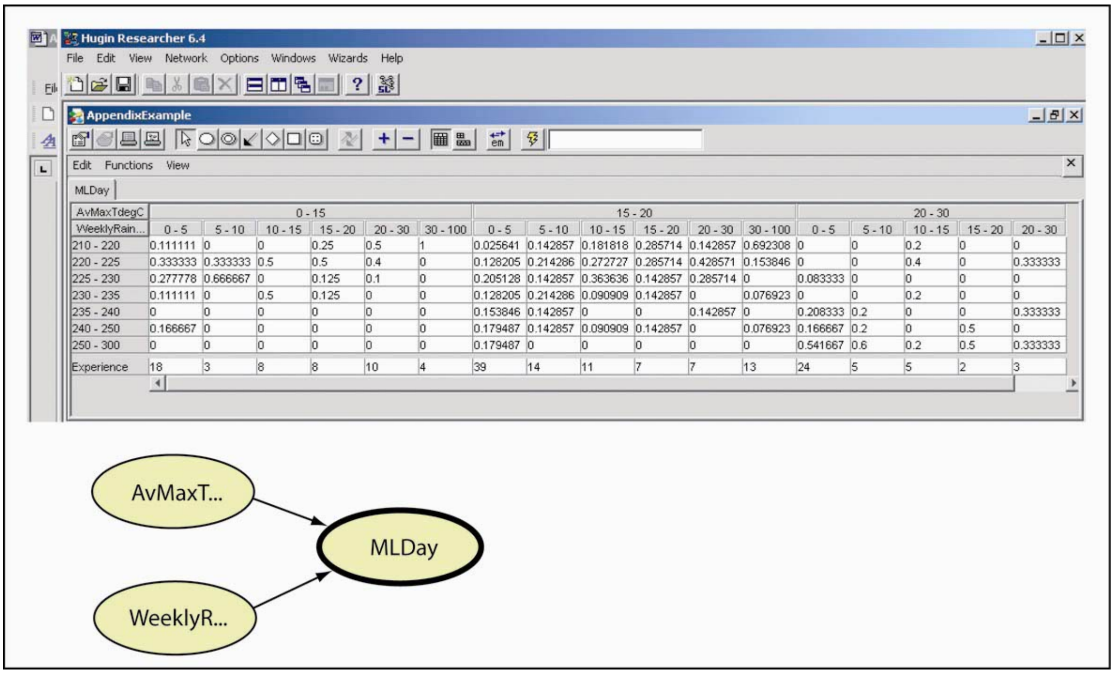
(b) Define data ranges (states)



(c) Initial distributions and experiences



(d) Run EM learning



AvMaxTdegC	0 - 15					
WeeklyRain...	0 - 5	5 - 10	10 - 15	15 - 20	20 - 30	30 - 100
210 - 220	0.111111	0	0	0.25	0.5	1
220 - 225	0.333333	0.333333	0.5	0.5	0.4	0
225 - 230	0.277778	0.666667	0	0.125	0.1	0
230 - 235	0.111111	0	0.5	0.125	0	0
235 - 240	0	0	0	0	0	0
240 - 250	0.166667	0	0	0	0	0
250 - 300	0	0	0	0	0	0
Experience	18	3	8	8	10	4

AvMaxTdegC	15 - 20					
WeeklyRain...	0 - 5	5 - 10	10 - 15	15 - 20	20 - 30	30 - 100
210 - 220	0.025641	0.142857	0.181818	0.285714	0.142857	0.692308
220 - 225	0.128205	0.214286	0.272727	0.285714	0.428571	0.153846
225 - 230	0.205128	0.142857	0.363636	0.142857	0.285714	0
230 - 235	0.128205	0.214286	0.090909	0.142857	0	0.076923
235 - 240	0.153846	0.142857	0	0	0.142857	0
240 - 250	0.179487	0.142857	0.090909	0.142857	0	0.076923
250 - 300	0.179487	0	0	0	0	0
Experience	39	14	11	7	7	13

AvMaxTdegC	20 - 30				
WeeklyRain...	0 - 5	5 - 10	10 - 15	15 - 20	20 - 30
210 - 220	0	0	0.2	0	0
220 - 225	0	0	0.4	0	0.333333
225 - 230	0.083333	0	0	0	0
230 - 235	0	0	0.2	0	0
235 - 240	0.208333	0.2	0	0	0.333333
240 - 250	0.166667	0.2	0	0.5	0
250 - 300	0.541667	0.6	0.2	0.5	0.333333
Experience	24	5	5	2	3

e) Conditional probability and experience tables for variable ML day^{-1} (Domestic consumption)

Figure 4.16 Screen dumps showing the main steps of the EM learning option

Description of steps shown in Figure 4.16:

Step (a) The first step selects the file containing the data needed for the CPT. In this case there are three variables ‘AvMaxTdegC’ (Max Weekly Temperature), ‘MLDay’ (Domestic consumption) and ‘Weekly RainMM’ (Weekly Rainfall). It is possible to preview the entire file before proceeding.

Step (b) Next, decide which variables to include in the analysis; in the example all three have been selected (the ticked boxes). At this point there is the opportunity to discretise the data into the intervals shown in Table 4.6. You are invited to enter the intervals for each variable. Once the discretisation is complete you can view the resulting file.

Step (c) Next, enter any prior distributions or experiences, if these are known. In this case prior knowledge of climatic factors or domestic consumption is unlikely, so the table is re-set to 1, indicating no prior knowledge. The experience table is also set to zero because we wish these to be extracted from the data.

Step (d) Finally, the learning operation can be run to construct the CPT for the child variable ‘Domestic consumption’. The accuracy with which you wish the calculation to be executed is defined by the level of significance, but for most cases the default value of .05 is more than adequate.

Step (e) When the run has been completed the CPT for the child variable can be viewed in the Graphical User Interface. Note that all the cases have been estimated and the 'experience' count for each case has been entered in the experience table at the bottom. Cases with 20 or more counts (e.g. AvMaxTdegC '15-20' and Weekly Rain '0-5') will have far less uncertainty attached to them than cases with less than 5 (e.g. AvMaxTdegC '0-15' and Weekly Rainfall '5-10'). Should any of the cases have zero experience the probability will be still be calculated, but clearly there would be considerable uncertainty attached to this result and any new data is likely to change it significantly.

A good explanation of the way in which probabilities are calculated from case data is given in Cain (2001). For more detailed mathematical explanations you are referred to Jensen (1996) or the help menus of both the Hugin and Netica software. A brief explanation is also given in Appendix II of these guidelines.

6.3 Review network and amend as necessary

Once the all the CPTs have been completed they should be carefully checked and reviewed with the stakeholders. Remember it is the values in the CPT that will determine the outcome of your networks, so it is vital that there is agreement on the validity of these figures.

Stage 6: Stakeholder involvement

At this stage the main involvement of stakeholders will be in circumstances where you have no reliable data for a particular link and the only source of information is either from local stakeholders or experts.

6.1 Input stakeholder / expert opinion, where necessary

The use of stakeholder and / or expert opinion should be a last resort when all other sources of potential data have been exhausted. No matter how knowledgeable or experienced an individual there is always the danger of bias or selective memory when opinions are being given. Inevitably this adds uncertainty to the network.

Where input from stakeholders is used, it is important to keep the CPT as small as possible. If the table requires less than about 15-20 entries it is probably just about possible to persuade the stakeholder to provide figures for each entry; more than this number and it becomes too tedious and complicated to do manually.

Sometimes the table may be unavoidably greater than 20 combinations, in which case you are faced with the problem of entering a large number of probabilities and of asking the stakeholder an equally large number of questions. To overcome this problem Cain (2001) suggested the use of an Elicited Probability Table (EPT) in which the stakeholder is asked to complete a small subset of the probabilities; the remainder are then calculated using interpolation factors derived from the stakeholders response. These are then used to complete the CPT.

An example of an EPT is given in Table 4.7. A child variable, 'Total income' has three parents, 'Crop yield', 'Market price of crop' and 'Forest products income'. For simplicity each variable has been assigned only two states 'high' (H) and 'low' (L).

Crop yield	Market price of crop	Forest products income	Total income	
			H	L
H	H	H	100	0
H	H	L	90	10
H	L	H	70	30
L	H	H	60	40
H	L	L		
L	H	L		
L	L	H		
L	L	L	0	100

Table 4.7 Example of an EPT (Elicited Probability Table)

You may decide that asking a farmer to enter probabilities for all 16 possible states of the child node is too confusing. What the EPT does, is to allow you to enter a subset of the probabilities and then to use this information to calculate the remaining values to complete the CPT. In this case the farmer has provided the probabilities for the two extreme situations and three others, leaving you with the task of calculating the remaining four. Details of the way in which the calculations are made are given in Appendix III and Cain (2001). A copy of Cain's EPT calculator is provided on the CD accompanying these guidelines.

4.8 STAGE 7: Check, collect feedback and evaluate network

Stage 7: Network construction

By this stage network design and data entry is complete. When the network has been compiled and the controlling or implementation (action or scenario defining) variables instantiated (i.e. the states selected by the user), the probability distributions for all the linked variables will change according to the actions/scenarios that have been set. There are now a few final steps to completion (Figure 4.17), these are:

- *Check consistency of the network*
- *Receive feedback from the stakeholders and if appropriate incorporate*
- *Add decision nodes if these are required*
- *Carry out sensitivity analysis on key variables*
- *If necessary link the Bn within other decision support systems*
- *Evaluate alternative options and make decision*

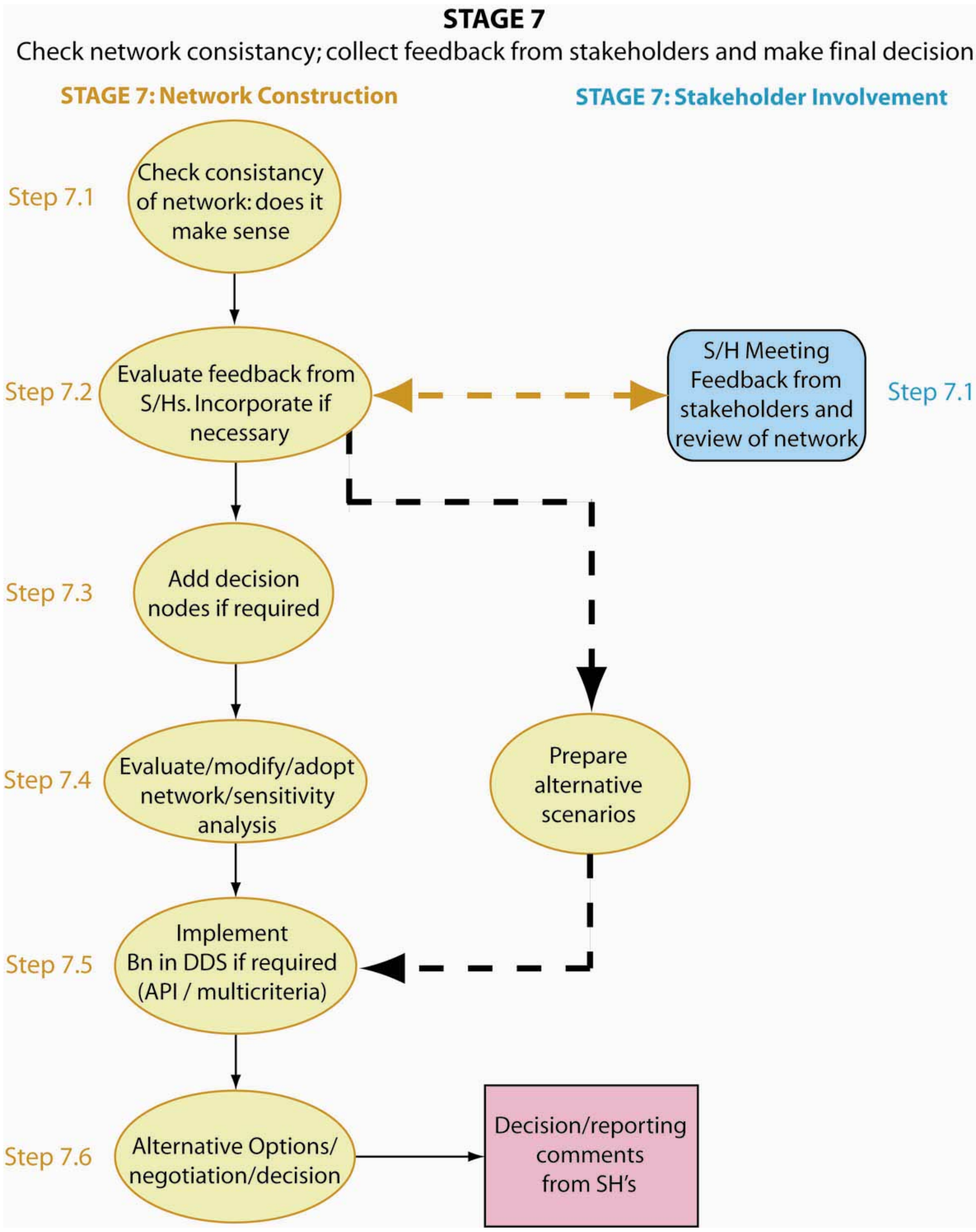


Figure 4.17 STAGE 7: check, collect feedback and evaluate network

7.1 Check consistency and logic of network

The first step is to check whether the network behaves in a logical way. This can be done by noting the impact of each implementation and controlling variable one by one. Set all implementation and controlling nodes to represent the current situation. Then change one of these to another state to represent a different action or scenario. Now note the impact on the variables linked to it and whether they respond in a logical way. For instance if a variable 'Rainfall' is changed to represent a wet year, the 'River flow' variable linked to it should show an increase. If not, and if no other variables are having an impact, it is likely the values in the CPT are wrong and need to be re-examined. Once the first variable has been verified re-set this to the current or base position and select another for checking in the same way. You should do this for all controlling and action or scenario variables in turn. Make a note of the impact of each variable on the objectives.

The impacts made on any indicator variable will be controlled by two factors:

- *The structure of the network:* the degree to which an indicator will change state depends on the number of intermediate nodes separating it from the implementation variable. The more intermediate variables separating the action from the indicator, the less the impact will be, because of the modifying effects of the intervening CPTs. It is also likely that the intermediate nodes themselves will have parents, adding to the dilution effect. The effect of this feature needs to be taken into account when results are being evaluated. As an example look at Figure 4.18 taken from the Spanish case study. The action variable here is named ‘APCON’ and the indicator ‘MB5’; what they are doesn’t matter. Note that even when extreme states of ‘APCON’ are selected, the impact on the objective variable (‘MB5’) is small. Also observe that the impact on each variable increases the closer it is to the action. This effect of structure needs to be carefully considered, and whether or not all of the intermediate nodes are necessary to describe the system. If they are they should be retained, but if you believe that some are not needed they ought to be removed and the network re-structured.

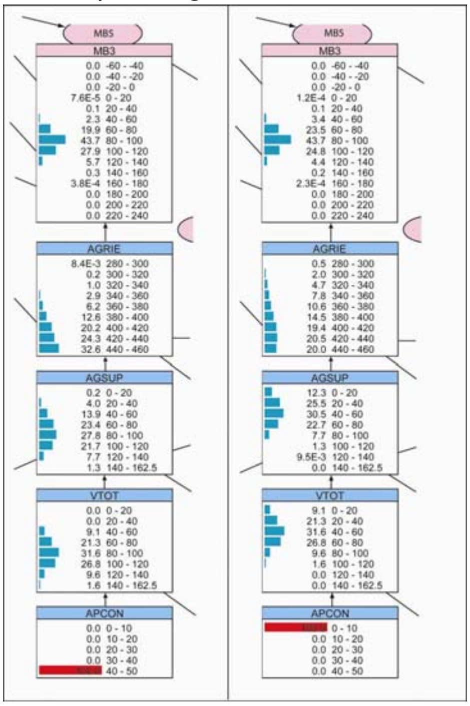


Figure 4.18 Modifying effect of intermediate variables

- *The values placed in the CPTs:* The strength of an action on an objective will be determined by the values placed in the CPT. When action is simulated the response in the objective should ideally result in a strong response. If the response is very subdued you should consider why this might be. There are 3 possibilities:

  (a) The structure and the influence of intermediate variables (Figure 4.18)

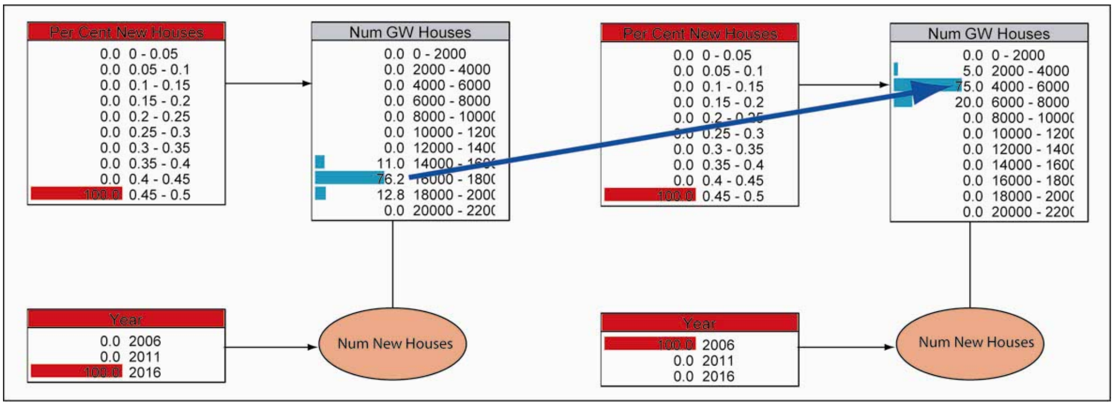
  (b) The impact of a second parent is cancelling out any effect

  (c) The values in the CPT are not sufficiently large

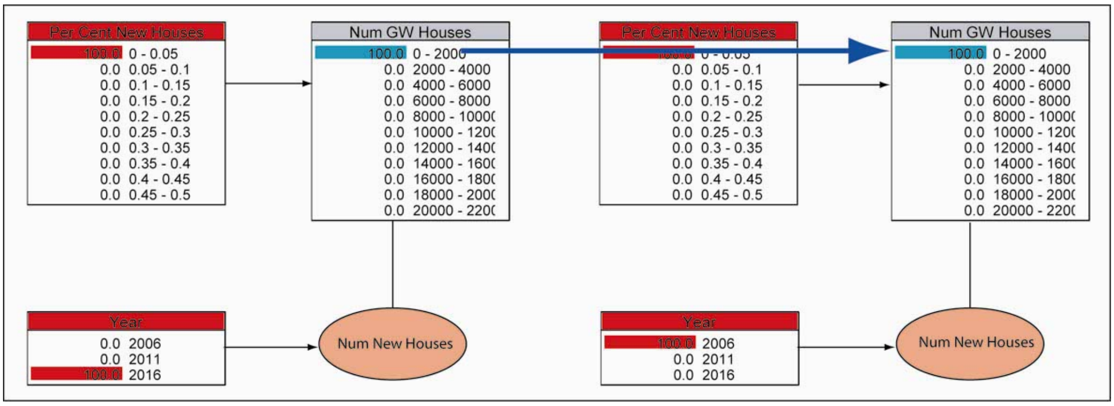
The first case (a) has already been discussed above.

To illustrate case (b), where a second parent has a modifying influence, an example from the UK network can be used (Figure 4.19). This predicts the number of new houses that may have a grey water recycling facility installed in the future (Num GW Houses). It depends on two variables; the period of time over which the prediction is to be made (Year) and the percentage of houses taking up recycling (Per Cent New Houses). The year selected controls the total number of new houses built; the further into the future, the greater the number of houses (Num New Houses). The percentage of these houses having grey water systems is set by the variable. Figure 4.19 (a) shows that when ‘Per Cent New Houses’ is set at a high value (0.45-0.5), changing the state of ‘Year’ has a significant impact on the child ‘Num GW Houses’.

However, when the same change to ‘Year’ is made, but with ‘Per Cent New Houses’ set at a much lower value (0-.05), the effect on ‘Num GW Houses’ is negligible (Figure 4.19 (b)). The effect of variable ‘Year’ is said to be *modified* by ‘Per Cent New Houses’. In this case the reason for the modification is clear (i.e. the percentage of new houses selected controls the total number of houses being evaluated). But sometimes the reason may be more subtle and not so immediately obvious.



(a)



(b)

Figure 4.19 Modifying effect of second parent. ‘Per Cent New Houses’ is a modifying parent to ‘Year’. In case (a) the effect of changing ‘Year’ on the child variable ‘Num GW Houses’ is strong. However, in case (b) changing the state of ‘Per Cent New Houses’ dramatically reduces the impact of ‘Year’ on ‘Num GW Houses’.

So long as the reason for the modifying effect is understood and recognised, then this feature is not a problem. On the other hand if it is not understood or logical, a re-structuring of the network should be considered.

The third case (c), where a subdued response is due directly to values in the CPT, and not to the effect of intermediate or modifying variables, should cause you to consider whether the link is really necessary. If not, it should be removed, unless of course you specifically wish to highlight the weakness of the link and bring it to the attention of stakeholders. Where an action results in a wide or regular distribution of probabilities in the objective, it indicates the information used to construct the CPTs is uncertain and should, if possible, be improved. Highlighting the presence of poor quality or absent data is a practical feature of networks that helps identify weaknesses in the understanding of a system.

Checks in diagnostic mode

A check of the network can also be made in what is known as diagnostic mode. Normally the state of the parent is set in order to predict the state of the child. But when in diagnostic mode the state of the child is fixed, which then propagates back to describe the states of the parents needed to generate that condition. In Figure 4.20, taken from the Danish network, the first case (a) shows that if the parent variable 'Land Use' is in the state 'Set Aside' – 60.37%, and 'Grass' – 39.63%, then the 'Hunting / fishing' child variable will be in the state 200 Dkr ha^{-1} – 100%. In case (b) the network is used in diagnostic mode. Here the state of the 'Hunting / fishing' child variable has been fixed in the 200 Dkr ha^{-1} state; this has then propagated back to indicate that the best combination of land use to achieve this income is 60% 'Set Aside' and 40% 'Grass', which confirms the original prediction. The diagnostic mode can be used to determine the land cover required to generate any selected level of income. These results of course assume that all other parents are in the same state throughout.

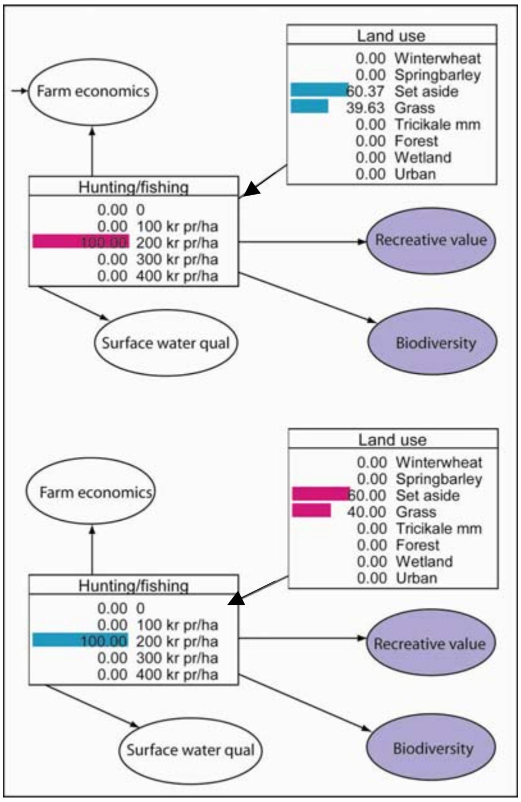


Figure 4.20 Network shown in (a) normal predictive mode and (b) diagnostic mode

7.2 Receive and incorporate feedback from the stakeholders

When the network is complete the results of different combinations of actions or scenarios on the objective variables can be presented to the stakeholder. This presentation should be done in an organised way so that stakeholders can note the effects of individual actions or scenarios, and then of different combinations. All results should be noted down for discussion. If anomalies or problems are observed they should be noted and the reasons identified. In some cases the structure of the network or values within CPTs may have to be modified, but this should only be done if new evidence, observations or insights are produced. Where new insights are gained it may then be necessary to introduce new variables, links or states. If stakeholders fail to agree on a particular aspect of the network one possibility is to produce two alternative solutions reflecting the opinions of the different groups. This was done for part of the Danish compensation network as illustrated in Figure 4.21.

In this instance the local farmers and environmental organisations could not agree on the extent to which pesticides pollute deep groundwater supplies. The farmers, being part of the problem, not surprisingly had a more optimistic view of the impact than the environmentalists. Instead of choosing one or the other, it was eventually decided to incorporate both opinions into the network for later evaluation.

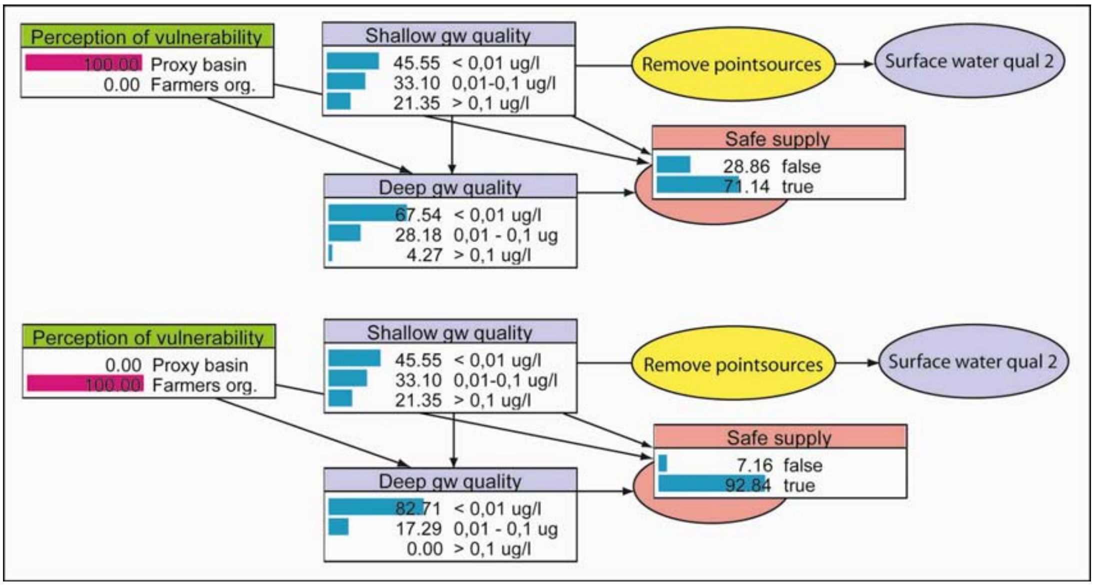


Figure 4.21 Taking account of different stakeholder opinions. The ‘Perception of Vulnerability’ variable has two states, one to represent the view of environmental organizations (Proxy basin) and one to represent the views of local farmers (Farmers org). You can see the farmers are more optimistic about the likelihood of a safe supply of water

7.3 Add decision nodes if required

So far these guidelines have assumed that the networks are being used for ‘probabilistic inference’; in other words they are being used to determine new beliefs in the light of observations that are made, or facts that are gathered. The network is composed entirely of what are known as ‘chance’ or ‘nature’ nodes. Used in this way a network does not make a decision, it simply presents the impacts of different actions or scenarios on specified objectives. It is up to the designer, in conjunction with stakeholders, to decide on the most appropriate action to take. In this sense the network simply informs. However, with the addition of extra nodes the network can be converted into a ‘decision network’ or ‘influence diagram’; examples are the Danish ‘compensation’ and Italian ‘environmental performance’ networks shown in Figures 4.22 (a) and (b).

Two extra types of node are required to convert a belief network to a decision network. The first are *‘decision nodes’*, traditionally drawn as rectangles. These are used to represent variables you are able to control. The second are *‘utility (or value) nodes’*, used to represent quantities you wish to maximise, and usually represented as a diamond or flattened hexagon.

The decision network functions by maximising the expected value for the utility node, thus ‘solving’ the problem. Links that go into a decision node act as *information links.* These links inform what is known at the time the decision is to be made. In other words the decision maker will know the values of all the nodes which have links to that decision node, but not the values of other nodes. The network will find a value for each configuration of the decision node parents; this is specified in the Utility node table. The example in Figure 4.22 (a) has a number of variables linked to the Utility node associated with the decision node ‘All Farm Economy’. The utility here is the amount of money the farmer is likely to obtain for different contract / payment schemes, given the states of the four variables linked to the Utility node (i.e. ‘Compensation’, ‘Animal / vegetable’, ‘Farm economics’ and ‘Land use’) . In the example we have assumed that the farmers have elected to grow spring barley as their main crop. Given this situation the decision node indicates the amount of money in DKr ha^{-1} that three different contract schemes are likely to provide. In this case the MJV[11] scheme offers the highest (6230 Dkr ha^{-1}), while no contract at all offers the lowest (3075 Dkr ha^{-1}).

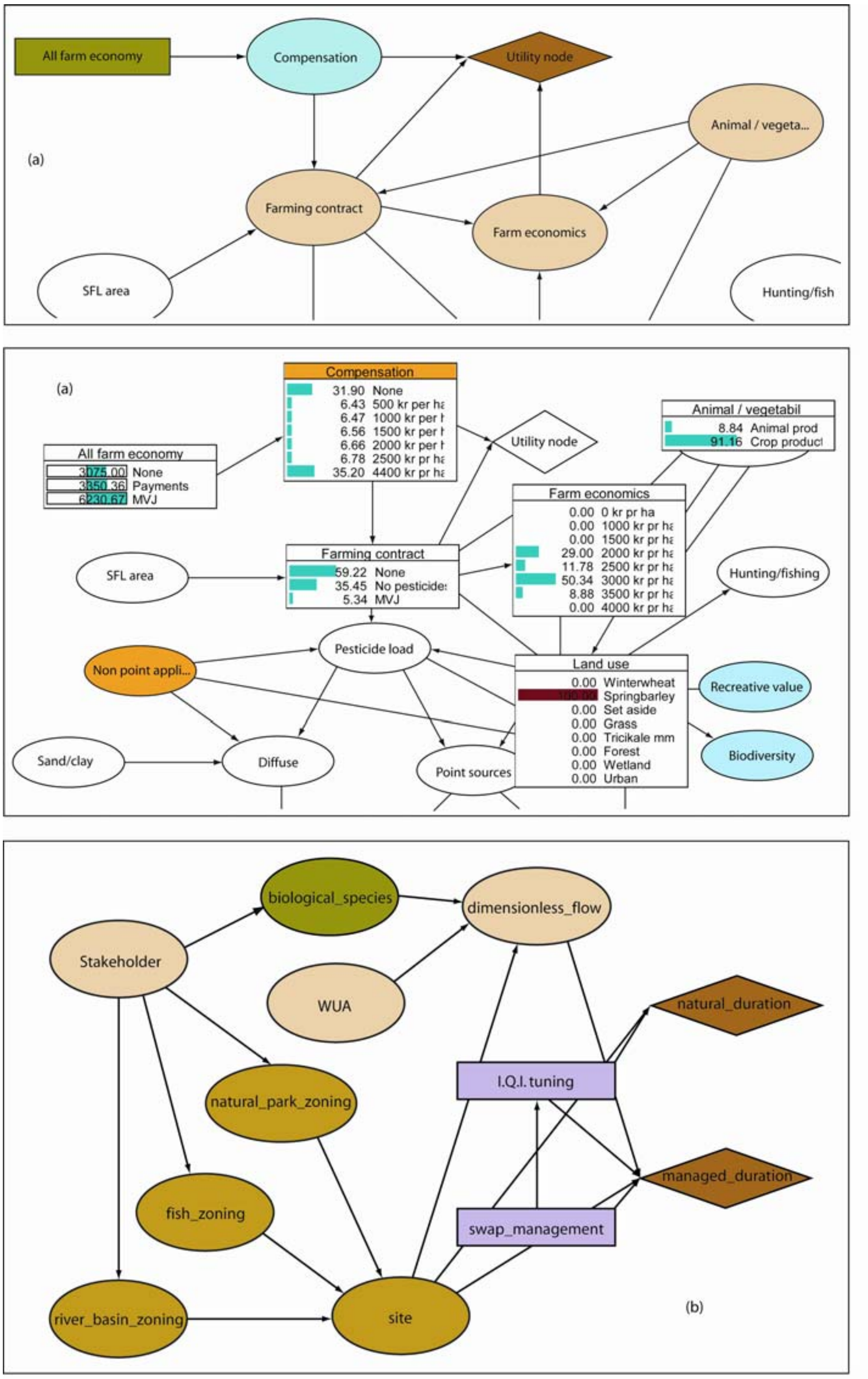


Figure 4.22 Decision and Utility nodes in the Danish 'compensation' network (a) and in the Italian (CESI) 'environmental performance' network (b)

[11] MVJ is the Danish abbreviation for 'Environmental friendly agricultural agreements'; these agreements demand zero pesticide application as well as imposing restrictions on the use of fertilizers and the choice of crop.

It is possible to have more than one decision node in a network. The Italian ‘environmental performance’ network shown in Figure 4.22 (b) has two decision nodes ‘Swap_Management’ and ‘I.Q.I tuning’. When more than one decision node is used it is necessary to ensure that if one decision depends on the result of another, they are arranged in the correct order. More information on influence diagrams and the rules of use can be found in the help files of software packages such as Hugin and Netica.

7.4 Network evaluation and sensitivity analysis of key variables

The completed network should now be evaluated to test the sensitivity of key variables under different conditions and its effectiveness as a decision making tool.

Sensitivity of key variables: Networks should be tested for their sensitivity to key variables in the same way and for the same reason that physical, economic and other types of models are tested. Sensitivity analysis can be used to identify which of a range of potential actions or scenarios are likely to be the most effective in dealing with a particular problem. It also helps to pinpoint places in the model where uncertainty can least be tolerated. A simple example is shown in Figure 4.23 taken from the Spanish network where the variable ‘ASUP’ (annual inputs from surface water) has a much greater impact on ‘VENT’ (water inputs) than ‘REC’ (recharge). In this instance it is clearly more beneficial to spend time and resources in providing better data input to ‘ASUP’ than ‘REC’ because of its much greater impact. The response of child variables to different states of their parents should be made in turn, to assess the degree of impact. Then different combinations of parent variables should be tested for their impact on the child. Each case should be noted and saved for consultation during the decision making process.

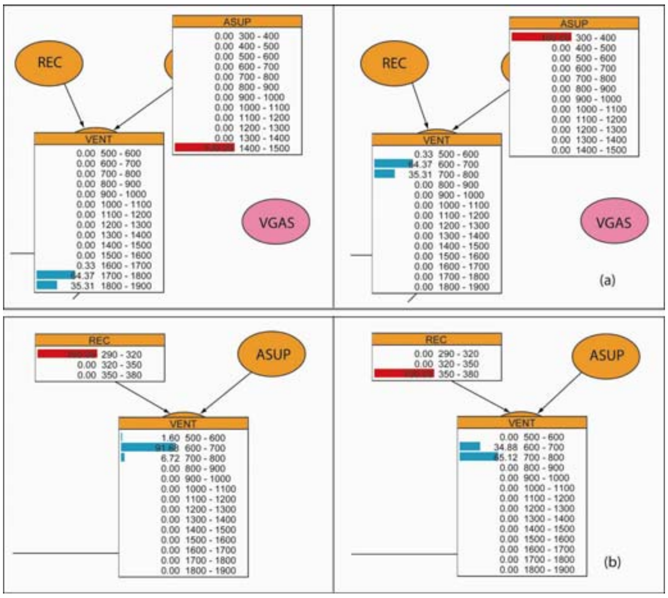


Figure 4.23 Evaluation of network; sensitivity analysis. Some actions or scenarios have a much greater impact than others on child variables. Here the variable ‘ASUP’ (a) has a much greater impact than ‘REC’ (b) on the child variable ‘VENT’

Effectiveness of network: The effectiveness of a network as an aid to decision-making depends to a large extent on the degree to which stakeholders regard it as a reliable means of reaching a fair and equitable decision. If stakeholders have been fully involved and consulted throughout its construction and a sense of ownership has emerged, there is a far better chance that decisions based on the network will be accepted. On the other hand, you cannot expect a similar degree of acceptance in cases where networks have been constructed by ‘experts’, with little or no reference to stakeholder opinion and knowledge. Even where involvement has been encouraged there is still no guarantee that unconditional agreement will be reached. This is demonstrated by the Danish networks where, despite extensive efforts to engage stakeholders throughout the entire process, feedback was not entirely positive. These comments are described in detail in the report ‘Test of Bayesian belief network and stakeholder involvement’, which is included on the attached CD.

7.5 Coupling of Bn to other models

In some circumstances your network may not be able to represent some aspects of the environmental system being investigated. This may happen, for example, when dealing with a process that involves calculation over multiple time steps. Bayesian networks are adept at coping with systems that are unstructured and where knowledge is limited, but where repeated calculations over successive time steps are required they have limitations. One way to overcome this problem is to couple the Bn to an operational model which is able to deal with the time varying aspects of the system. In these circumstances the Bn is used to represent the less structured part of the system, from which information is then fed into the operational model. Once model calculations have been made, the output information is returned to the Bn and the process repeated. To do this requires an understanding of the application programme interfaces (APIs) used by software packages such as Hugin and Netica; but this is outside the scope of these guidelines. However, it is useful to know that the facility to couple networks to, or embed them within, models can be undertaken given sufficient understanding of the software packages.

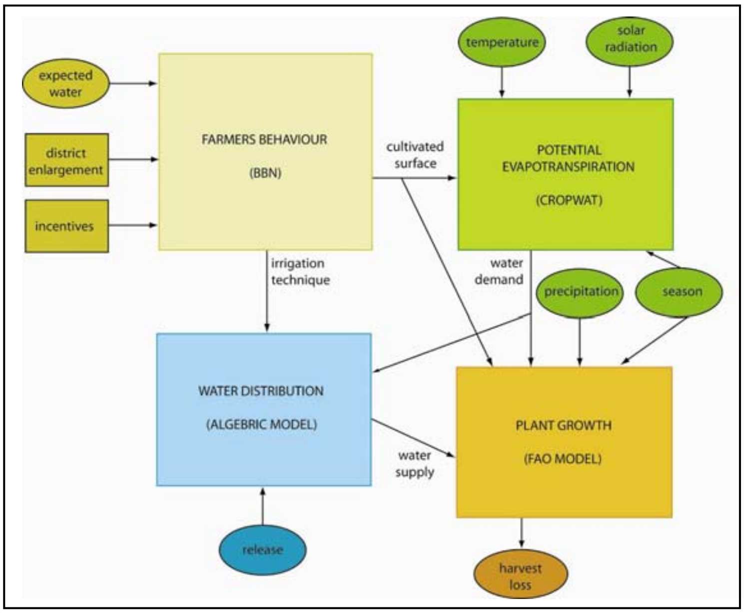


Figure 4.24 The relationship of the Bn to the irrigation district models

An example of this type of application is provided by the Vomano River ‘reservoir network’ (Figure 4.24). In this study, flow in the Vomano is controlled by a series of 3 reservoirs for the generation of hydroelectric power. These releases are controlled by the electricity generating company, and scheduled to suit the demands of the electricity grid. But there is also a demand from farmers for irrigation water at the lower end of the valley. The

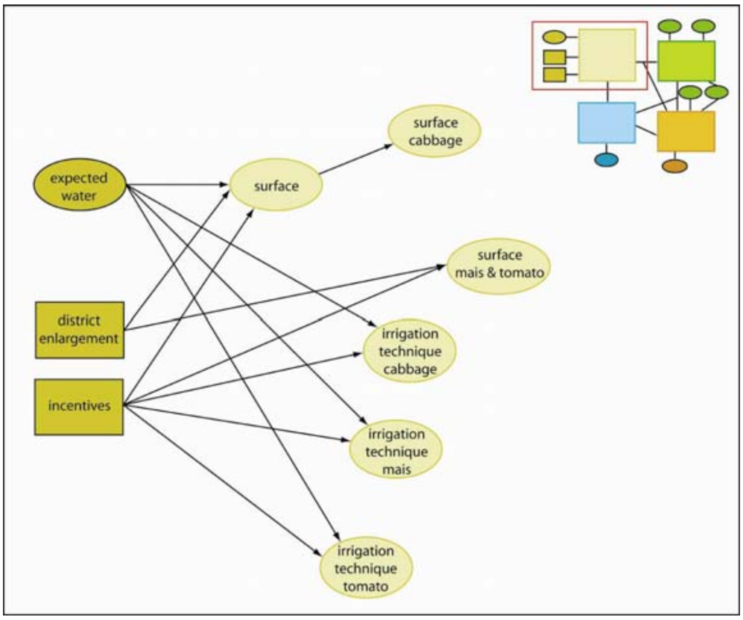


Figure 4.25 The Bn used to simulate farmer behaviour

problem is to optimise water releases to the advantage of both the farmers and the electricity company.

Currently all releases are scheduled using an operational model of the reservoir system, with no account taken of irrigation requirements. The demand for irrigation water depends not only on the area under cultivation, but on a number of other factors including the efficiency of the application systems being used. In turn, the efficiency of these systems partly depends on the scale of incentives and subsidies and the behaviour and attitude of the farmers themselves. This type of problem cannot be incorporated within the existing operational model, which simply calculates water transfers, but it can be addressed using a Bayesian network. A network to simulate farmer behaviour in the region has been developed and linked to a series of small models used to help calculate water demands. This is illustrated by Figure 4.24. Details of the Bn itself are shown in Figure 4.25. By coupling this network and its associated models to the operational reservoir model it has been possible to balance the two demands. Details of the network, the way it has been linked to the operational reservoir system and the results are given on the attached CD.

7.6 Weigh up alternative options and make decision

Once the final network has been completed and the full range of management options and / or scenarios run, a final decision can be made taking into account the various criteria for success and failure that will have been agreed earlier. However, you are still likely to have some disagreement between stakeholders. Under these circumstances a decision based on available evidence needs to be made, and the reasons for your decision presented to the stakeholders. To do this the Bn can be used to quickly compare and contrast different options. If the Bn is used to explain the reasons for your decision, even though some may disagree with your final choice, you have at least made the process as transparent and open as possible and allowed the stakeholders full participation in the construction of the decision making tool.

In cases where the choice is difficult, you may wish to enter decision nodes to determine the utility of different choices. This may make the selection more straightforward, assuming the stakeholders agree with the choice of utility on which you are basing your decision.

Finally you may wish to use another technique to make your final choice. For example, the states of key variables under the full range of management options can be analysed using other types of decision support tool such as Multi Criteria Analysis (MCA). This may be a more acceptable way to make a final decision, but again it depends very much on the attitude and opinions of the stakeholder group.

Stage 7: Stakeholder involvement

The final stage of the decision making process is to run through the range of management options, noting the impact of each option on the indicator variables. Stakeholders should be given the opportunity to comment on the results and to include extra options if these are deemed to be relevant. This is a crucial step in the process since any decision that is endorsed will carry more authority with stakeholder support.

7.1 Feedback from stakeholders and review of network

The impact of individual implementations on key indicator variables should be noted before combinations of parents are investigated. An evaluation of each run can then be made based on criteria agreed with the stakeholders.

For a network that contains only one or two indicators and actions the decision may be quite easy to make; however, where multiple implementation and indicator variables are involved, the problem is more complex. With complex networks the best approach is to run a range of different implementations to assess which have the biggest impact. Those having most impact can be given more weight. It is also important to evaluate the effects of different combinations of actions / scenarios. A combination of implementations may achieve the objective in a more acceptable way than simply selecting a single option.

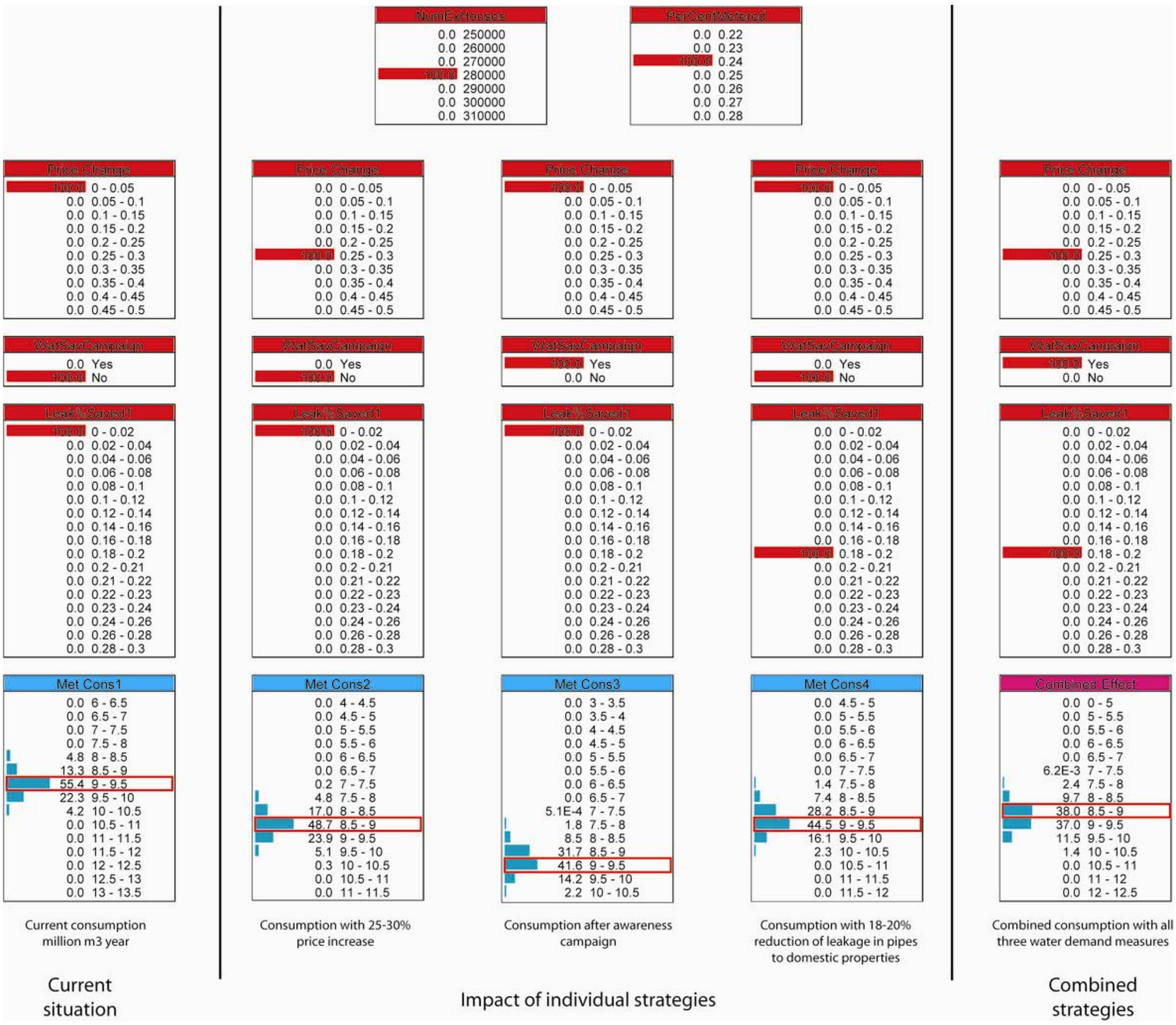


Figure 4.26 Network output from the UK 'Water Demand' case study. The effect of different water saving measures on water consumption in metered households is shown. The top row indicates the number of houses that are currently metered. The next three rows show the intervention actions; the bottom row is the resulting total consumption in $Mm^3\ yr^{-1}$. Present day consumption is represented in the first column. The individual impacts of price changes, awareness campaigns and leakage reduction are given in the next three columns. The combined effect of all three measures is given in the right hand column.

To illustrate the way in which network outputs can be used to focus discussion before final decisions are made, a selection of output examples from each of the MERIT case studies can be used. The first (Figure 4.26) is taken from the UK 'Water Demand' network. This example shows the effect of different water saving measures on water consumption in metered households. The number and percentage of houses metered in the region is shown in the two boxes on the top line. The current situation is presented in the left hand column, where actions are shown in the top three windows and the resulting total consumption (millions m^3 yr^{-1}) in the bottom window. The three columns to the right of the current situation represent the impacts of implementing single actions; price increase (column 2), implementing water saving campaigns (3), and improving leakage control (4). Column 5 represents a combination of implementations.

It is clear that the most effective measure for reducing consumption is price change, but only in houses fitted with water meters. A 25-30% annual price increase results in overall reduction in demand of 6%. This is obtained by comparing the consumption figures for the current situation (column 1) against the figures after the price change (column 2). You can see that before the price change, the most probable consumption (55.4%) is 9 to 9.5 Mm^3 yr^{-1}; after the price change the most likely range of consumption (48.7%) is reduced to 8.5 to 9 Mm^3 yr^{-1}.

In contrast the effect of introducing water saving campaigns (column 3) and reducing the level of leakage of service pipes to houses (column 4), is less than 3%. The implication is that money spent on awareness campaigns and leakage reduction will be less effective than a political initiative to authorise price increases. Combining all 3 scenarios (column 5) gives a reduction of about 9%, but of course there are cost implications involved with this option. The network can be used by the operator to decide whether or not the expense involved with leakage reduction and promoting awareness campaigns is merited by the resulting decrease in consumption. It also provides evidence to support a political initiative to authorise significant price increases.

Another example, taken from the Danish 'flooding' network, is shown in Figure 4.27. Here the impact of two wetland creation projects is evaluated with respect to the current situation. The actions taken are indicated by the red bar in the sixth row of windows; the key indicator variable, Wetland (2), is shown in the top row. The current situation is given in column 1, the impact of project 1 is in column 2, project 2 appears in column 3, and in column 4 the impact of project 2 is also shown, but using a different data set for the historical record variable.

The more successful an action, the shallower the water table, and more water-logged the wetland becomes. Using this criterion, the most effective option is project 2 which has a significant effect in raising water levels in the area of planned wetland (2). Project 1 on the other hand has little impact. It is worth pointing out that the results differ slightly depending on the set of data used to build the network. Conditions at wetland (2) are marginally better when a dataset for the period 1998-2002 is used; this is a timely reminder that, like any model, the quality of the network is dependent on the quality of data on which it is based.

Discussion of this network with the Danish citizen stakeholder group led to the recommendation that more variables need to be added to describe the impact of abstraction of existing wellfields on regional groundwater levels. It was also recognized that some validation of the network using either new data, which currently is not available, hydrological model output, or remote sensing information would be advisable before final decisions were made.

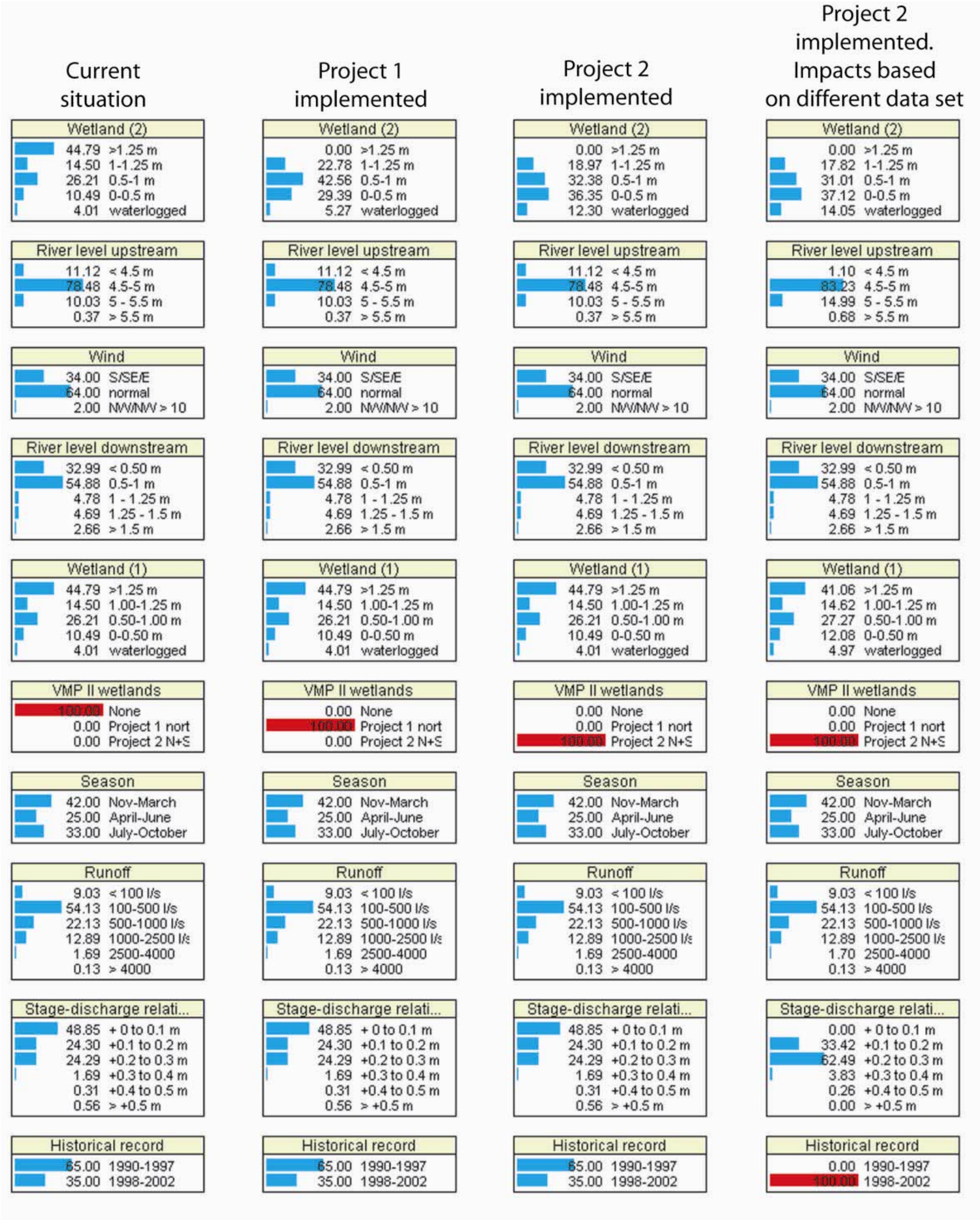


Figure 4.27 The Danish ‘flooding’ network showing the results of different scenarios for the establishment of future wetlands in the Havelse catchment

In Figure 4.28 part of the output from the Spanish ‘irrigation’ network is shown. This part of the network deals with the variables influencing the use of water for irrigation. The variables include: the type and distribution of crops (‘Crops’); the volume of water available for irrigation (‘Irrig Water’), the volume of available surface water (‘Surface water’), the income of farmers in the La Mancha region (‘Income’) and the sustainability of the east La Mancha aquifer (‘Aquif2’).

The example shows the way in which farmers' income can be increased by changing the distribution of crops, but at the expense of increasing abstraction for irrigation,

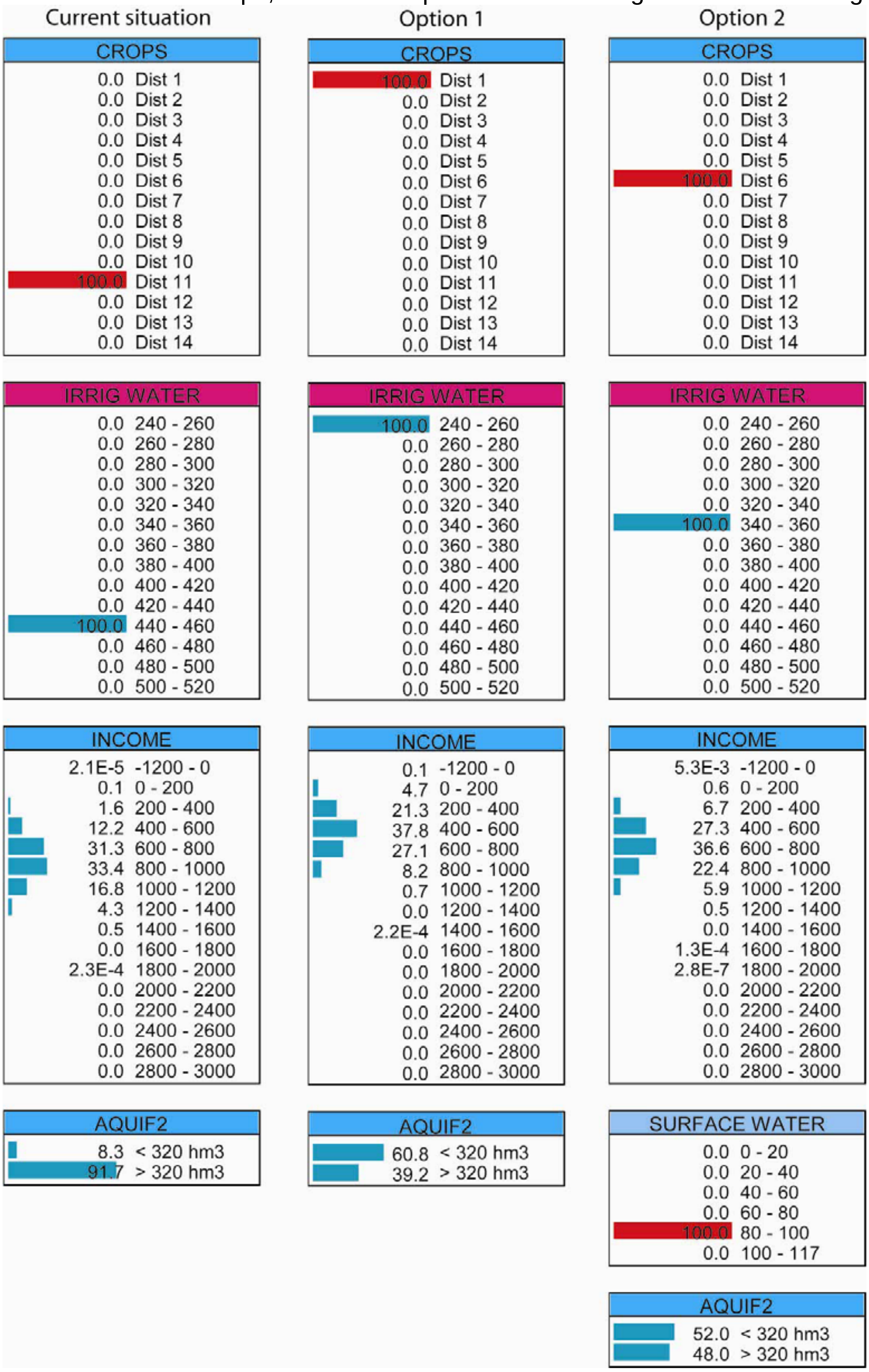


Figure 4.28 Part of the Spanish network showing the trade off between Farmers' income and Aquifer sustainability. The sustainable discharge from the aquifer is 320 $hm^3\ yr^{-1}$.

thereby reducing the sustainable status of the aquifer. Practically all irrigation water in the region is supplied from groundwater sources. The sustainable discharge from the aquifer has previously been calculated by a number of studies to be 320hm^3 yr^{-1}. The left hand column shows the current situation, which indicates a 91.6% probability of the present discharge rate from the aquifer being unsustainable; a disastrous situation. The scenario in the second column (Option 1) is more promising and shows a 60.8% probability of a sustainable condition being maintained. The trade off, however, is that farmers' income is reduced from 800-1000 Euros ha^{-1} to 400-600 Euros ha^{-1}. The third column (Option 2) shows a compromise solution which maximizes farmers' income, but at the same time ensures aquifer sustainability. This is done by augmenting irrigation supplies from surface water sources, indicated through the introduction of the 'Surface Water' variable. In this case farmers' income is maintained at 600-800 Euros ha^{-1}, while the probability of the aquifer being sustainable is 52%. This option, although achieving the aims of planners has a financial cost, which needs to be taken into account before a final decision is made. The option is one of several currently being considered.

The presentation and subsequent review of results for the Italian 'reservoir' network was different, because in this case the network was linked to an operational management system designed to regulate flows between reservoirs along the Vomano River. Linking the network to the operational model meant sacrificing the graphical facility of the network package so it was not possible to present the output of simulations in the conventional way. Instead the network designer was forced to present the results of different scenarios in the form of a matrix shown in Figure 4.29. This matrix identifies the impact of a number of alternative scenarios on a range of indicators chosen by the stakeholder group. The basic issue was to ensure a fair allocation of water between hydropower generation and that needed for irrigation at the downstream end of the valley. There was also the question of whether the water resources were sufficient to allow a doubling of the present irrigated area.

A large number of alternative options were considered, but the two extremes were (1) maximizing hydropower revenue, or (2) minimizing agricultural losses, both at the expense of each other. The results demonstrated that a more efficient daily operation of the reservoir system could significantly increase water availability for the irrigation district both under present conditions and in the case of an enlargement of the irrigated area. On the other hand a doubling of the pumping capacity to increase hydropower provides only a marginal improvement compared to the financial investment required for such an undertaking.

IMPACT MATRIX		INDICATORS							
		Hydropower		Irrigation			Environment	Potable water	MEF
		Income	Production	Deficit [Mmc/y]		Ricavo S.Lucia	yearly lacking flow	Piaganini uptake	
		[MIL Euro/y]	[Gwh/y]	on demand	on capacity	[MIL Euro/y]	[mc/s]	[Mmc/y]	
ALTERNATIVES	A0 History	34.77	527.65	16.16	12.35	0.43	0.50	3.74	1.00
	A1 Current situation								
	a. max. hydropower revenue	46.46	580.76	19.95	16.14	0.51	0.50	3.73	1.00
	d. min agricoltural loss lex to hydro	46.45	614.22	3.81	0.00	0.55	0.50	3.74	1.00
	c. compromise policy	47.33	620.19	3.81	0.00	0.53	0.50	3.74	1.00
	c.1 MEF interceptors								
	c.1.1 interceptor 1100 (30%)	44.76	580.88	3.83	0.02	0.51	0.50	3.57	1.00
	c.1.2 interceptor 400 (30%)	44.52	579.30	3.84	0.03	0.50	0.50	5.02	1.00
	b. max hydropower rev. Lex agr.	47.43	605.02	11.89	8.08	0.51	0.50	3.73	1.00
	b.1 Ruzzo uptake increase								
	b.1.1 a 1.4 mc/sec	46.61	590.25	12.56	8.75	0.50	0.50	13.76	1.00
	b.1.2 a 2.1 mc/sec	46.22	585.95	12.91	9.10	0.50	0.50	27.06	1.00
	b.2 Fucino river MEF								
	b.2.1 10% mean flow	47.29	602.07	11.85	8.04	0.51	0.45	3.73	0.90
	b.2.3 30% mean flow	46.96	595.86	11.74	7.93	0.51	0.35	3.73	0.70
	b.3 Mix of b.2.4 and b.1.2								
	b.3.1 2.1 mc/sec + 30%	46.55	583.32	14.71	10.90	0.50	0.35	19.26	0.70
	e. min agricultural loss	21.90	448.65	3.81	0.00	0.52	0.50	3.73	1.00
	A2 2 X irrigation district								
	a. max. hydropower revenue	46.46	580.76	44.84	44.84	1.04	0.50	3.73	1.00
	d. min agricoltural loss lex to hydro	40.97	592.94	0.03	0.03	0.97	0.50	3.66	1.00
	c. compromise policy	47.55	644.38	0.96	0.96	1.01	0.50	3.72	1.00
	b. max hydropower rev. Lex agr.	47.99	615.87	24.12	24.12	1.02	0.50	3.73	1.00
	e. min agricultural loss	22.20	460.54	0.00	0.00	0.87	0.50	3.73	1.00
	A3 2 X pumping								
	a. max. hydropower revenue	48.19	602.20	25.46	21.65	0.49	0.47	1.10	1.00
	d. min agricoltural loss lex to hydro	46.49	607.64	3.81	0.00	0.53	0.50	3.42	1.00
	c. compromise policy	47.17	604.44	3.81	0.00	0.50	0.47	3.42	1.00
	b. max hydropower rev. Lex agr.	47.50	606.21	11.89	8.08	0.51	0.50	3.73	1.00
	e. min agricultural loss	20.30	433.56	4.15	0.34	0.49	0.50	2.93	1.00

Figure 4.29 The impact matrix from the Italian ‘irrigation’ network derived from the output of the linked Bayesian network – reservoir operations model. The alternative scenarios are listed in the left hand column; the indicators in the columns to the right

Stakeholders were surprised at the outcome, because it showed that the current problems are not caused by a water shortage, but a lack of efficiency in the current system of water pumping. One of the compromise scenarios allowed a simultaneous increase in hydropower revenue and an increase in the area of irrigated land. This involved pumping water back up the reservoir system during the night, when demand for power is low. This particular example shows the power of linking a Bayesian network, which is able to take into account social and economic factors such as farmer behaviour and income, with a strictly quantitative operations model running on a daily time step.

Finally, Figure 4.30 shows some results from the Italian ‘environmental’ network, which incorporates decision nodes to evaluate the environmental cost of two different management policies. In Figure 4.30 (a) the different ‘environmental costs’ of the two water policies from the point of view of one of the stakeholders, the Gran Sasso national park, are identified. These costs are given as indices; the lower the index, the more environmentally attractive. From the results, policy 2 (10.33) is thus more environmentally attractive than policy 1 (10.69), at least from the national park point of view. The numbers in the ‘site’ node, represent the importance given by the national park to various locations along the valley, from where minimum instream flows will be released.

In the second diagram, Figure 4.30 (b), the ‘environmental cost’ from the point of view of a second stakeholder, the Provincia di Teramo (a local authority), is shown. Again policy 2 (10.9) is better than policy 1 (11.0), but at a higher cost than that perceived by the Gran Sasso national park. However, the importance attached to the various water release locations in the ‘site’ node is very different, reflecting the contrasting viewpoints of the two stakeholder groups.

The third Figure (c) shows the network being used in a ‘diagnostic’ mode, in which the state of the child is fixed and the states of the parent(s) needed to produce that condition, are propagated. In this example the state of the site node is fixed at Ponte Vomano; this means all the instream flow takes place from this point. Given this situation the ‘stakeholder’ node provides a profile of the type of stakeholder who would most prefer that condition. For this example there is an 82% probability that the stakeholder most likely to be in favour of the simulated condition is the Provincia di Teramo, with a 0% probability of acceptance by the Gran Sasso national park.

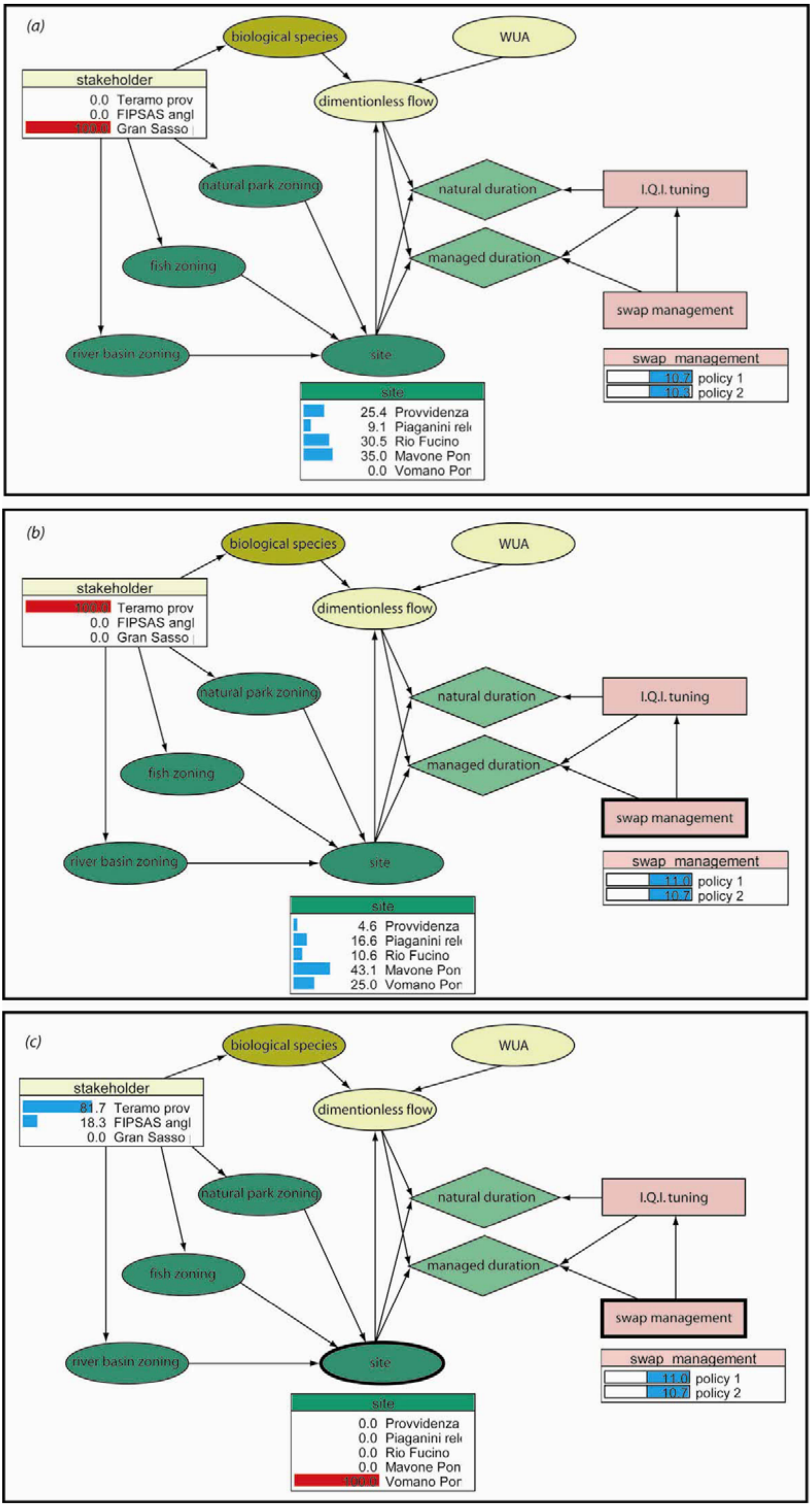


Figure 4.30 Different 'environmental costs' of two water policies from the point of view of (a) the national park; (b) the Provincia di Teramo; and (c) a generic stakeholder preferring all instream flow located at Ponte Vomano

5. MERIT: Case studies[12]

5.1 Water Demand network: United Kingdom

Background

In 2001, the UK Environment Agency (EA) launched a programme to improve the management of ground and surface water abstraction licensing in 126 catchments throughout England and Wales. For each catchment a Catchment Abstraction Management Strategy (CAMS) has been, or will be, developed in close consultation with local stakeholder groups. In future all abstraction licences will be periodically reviewed in the light of CAMS assessments. MERIT collaborated with the EA during the development of the CAMS for the Loddon, a catchment in South East England (Figure 5.1). The aim was to evaluate the effectiveness of Bayesian networks as an integrating technique, and the extent to which they can be used to encourage effective engagement of local stakeholders in a decision making process.

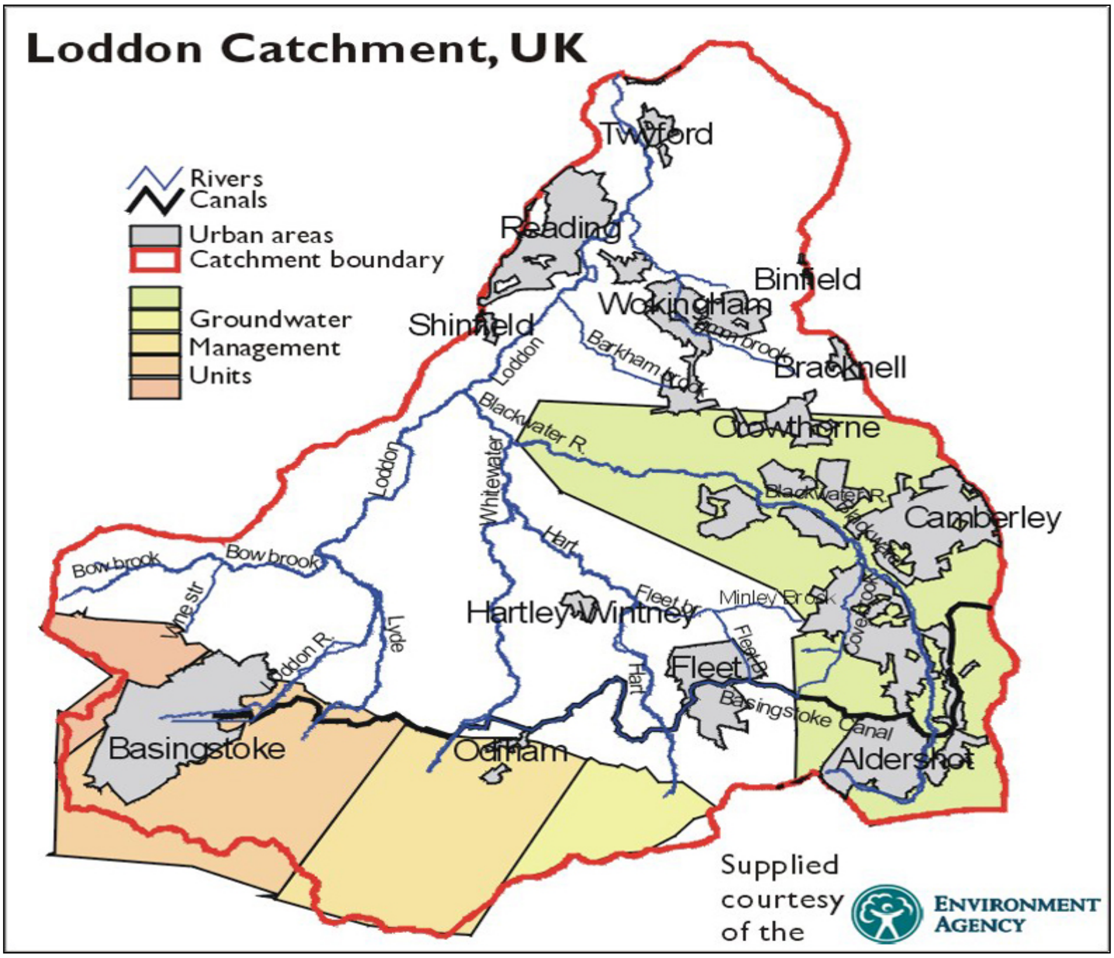


Figure 5.1 The Loddon Catchment

[12] Working copies of the UK, Spanish, Danish and Italian networks are included on the attached CD

Stakeholder consultation

Initial contact and selection of a representative stakeholder group followed procedures established by the EA. Stakeholder selection was based on the returns from 400 widely targeted leaflets explaining the aims of CAMS and inviting comments and expressions of interest. The selected group is shown in Table 5.1.

Organisation	Stakeholder responsibility
UK Environment Agency	Water resource management
South East Water	Strategic abstraction
National Farmers Union	Non-strategic abstraction
Hampshire and Isle of Wight Wildlife Trust	Conservation
Blackwater Valley Recreation Management Service	Recreation
Fisheries Consultative Committee	Fisheries
Planning Department, Wokingham District Council	Local Councils

Table 5.1 Loddon CAMS stakeholder group

The group comprised 7 national and local organisations believed to be representative of all major interests in the area. An independent facilitator, with no links to the Loddon, was appointed to chair all meetings. During the course of consultation the group met formally on 3 occasions following established EA procedures. The first meeting was used to inform the group of the aims of both the CAMS process and the MERIT project, and to allow each stakeholder to identify the water resource-related issues of most relevance to them, be this environmental, economic or social. A list of all issues was recorded; the three most notable were:

(a) The need to meet current and future domestic water demand in the region (Water Company)
(b) The need to ensure environmental allocation of water (Wildlife Trust)
(c) Protection of fisheries (Fisheries group)

At a second meeting the technique used to calculate available ground and surface water resources, on which the licensing decisions are based, was explained. An open and transparent explanation of the system was felt to be necessary to help gain the confidence of the group. An explanation was also made of the Bayesian network approach, with some examples given to highlight the potential of the technique. Finally, stakeholders were invited to raise any further issues that may have occurred to them since the previous meeting.

A third stakeholder group meeting was used to present the results of the resource assessment of the catchment. Despite a generally favourable assessment, one of the sub catchments (the Whitewater River) appeared to be over-abstracted and potentially required some remedial action. However, ecological, groundwater and gauged flow data have yet to show any negative trends, which conflicted with the initial resource assessment results.

Based on the results of the assessment, four management options were discussed with the stakeholder group:

- To maintain the current licensing policy (i.e. winter surface abstraction allowed, but no summer abstraction)

- To allow no more surface water abstraction
- To prevent surface abstraction when flow is below 37.8 Ml day^{-1}
- To improve the distribution efficiency and demand management in the area

An opportunity was given for all stakeholders to comment on these options. Essentially, all were prepared to agree with decisions taken during this implementation of CAMS because there was no direct impact on any of the interests represented, and little change in the licensing situation was likely to take place.

The initial plan was to use a Bayesian network to help resolve the selection of the different options for resource management across the region, using the technique to compare the pros and cons of each approach. However, the lack of conflict between each option meant that this comparative exercise would have been purely academic. Instead it was decided to construct a network to investigate just one of the options and use it to evaluate the way in which it might be implemented. The option selected was the improvement of water demand management in the catchment. The aim of the network was to investigate the impact of actions available to reduce the demand for domestic water.

The initial variables for the network were obtained during the first stakeholder meeting. However, to obtain more information individual meetings with the relevant stakeholders, South East Water and the Wokingham District Council, were arranged.

Water Demand Management Network

The demand management network was developed at the catchment scale. However, water companies compile supply and demand data according to 'Resource Zones'. In the case of the Loddon the catchment is covered mainly by Resource Zone 4 (RZ4) of the South East Water Company (Fig 5.2). The correspondence between the catchment and resource zone was sufficiently close to enable the data to be easily adjusted to the catchment scale, this being considered more relevant to the CAMS process.

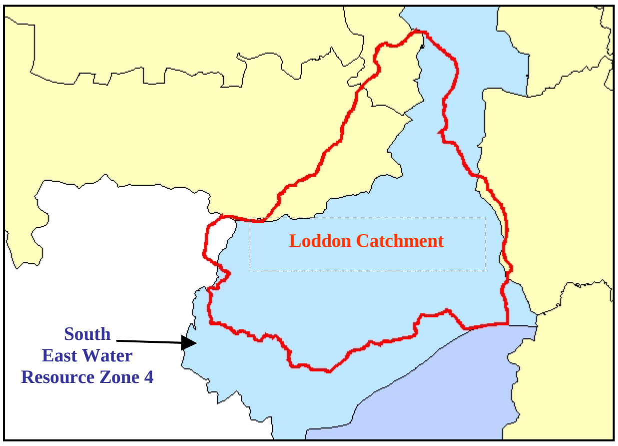


Figure 5.2 The Loddon catchment and Resource Zone 4 (RZ 4)

Four potential strategies to control domestic water demand were identified by stakeholders:

- <u>Pricing</u>: varying the price of water
- <u>Public Awareness</u>: campaigns making the public realise the importance of reducing demand
- <u>Grey Water Re-use</u>: the use of water from sinks, baths etc. for toilet flushing and for use in the garden.
- <u>Leak reduction</u>: reducing the amount of water lost through leaks (on private property, not main pipelines)

Data was collected mainly from the reports and Web pages of the water industry regulator Ofwat, the national office of statistics, the South East Water Company, Wokingham County Council and various publications. Other information was obtained during interviews with Water Company and Council staff.

Information collected from the water company and Ofwat was based on the resource zone area, which is slightly larger than the catchment area. Figures were adjusted to account for this difference. The total number of houses was derived from a GIS study of post code areas within the catchment boundary.

The network was designed to enable the effects of each of the four potential strategies to be identified individually (Figure 5.3). It was also designed to separate the impacts on houses that have water meters from those with no meter. The effect of a price change on a house with a meter, where the owner pays per m^3 consumed, will be different from a house without a meter where there is a set annual charge, regardless of how much water is consumed. Water consumption is expressed in millions of m^3 per year. There is also a distinction made between the existing housing stock and new houses planned for the next 15 years.

A feature of the network is that the links between the four strategies and water demand carry different degrees of certainty. The links for 'leakage' and 'grey water re-use' are based on actual data and are, therefore, well founded with a limited degree of uncertainty attached to them. On the other hand the effect of 'awareness campaigns' and 'pricing' are based on more qualitative information, so the associated uncertainty with these links is far greater. This identifies one of the strengths of Bns; the ability to link variables using whatever quality of information is available.

Results

The results showed that, so far as the current housing stock is concerned, the most effective water demand management tool available is price change, but only on those houses that are metered. With a 25-30% price increase the indications are that demand would reduce by up to 6% (0.55 million m^3) in metered houses; increasing the price by 45-50% reduces demand by 10% (0.92 million m^3). The impact of price on non-metered housing is insignificant, since in these cases water charges are not based on consumption, but a fixed annual fee. Unfortunately for the water company only 24% of housing stock is currently metered, so for pricing to become more effective this percentage needs to be increased.

The second strategy, raising the awareness of the public about water conservation, seems to have a limited impact, reducing demand by only 2% (0.2 million m^3), though it must be stressed that this is based on uncertain data. Nonetheless, the impact remains very small and as such the cost of this type of campaign needs to be carefully considered and set against the potential gains.

The final strategy to reduce demand in the current housing stock is the programme of leakage reduction from service pipes to domestic houses. Currently it is estimated that each property loses about 40 litres day^{-1}. The target for the water company is to reduce this leakage by 30% (i.e. to 28 litres per day). Reduction below this rate is not considered to be economic. With this maximum amount of leakage reduction, demand will be lowered by 3.5% in metered houses, and 2% in those that are unmetered. Overall this will account for a saving of 0.9 million m^3 $year^{-1}$.

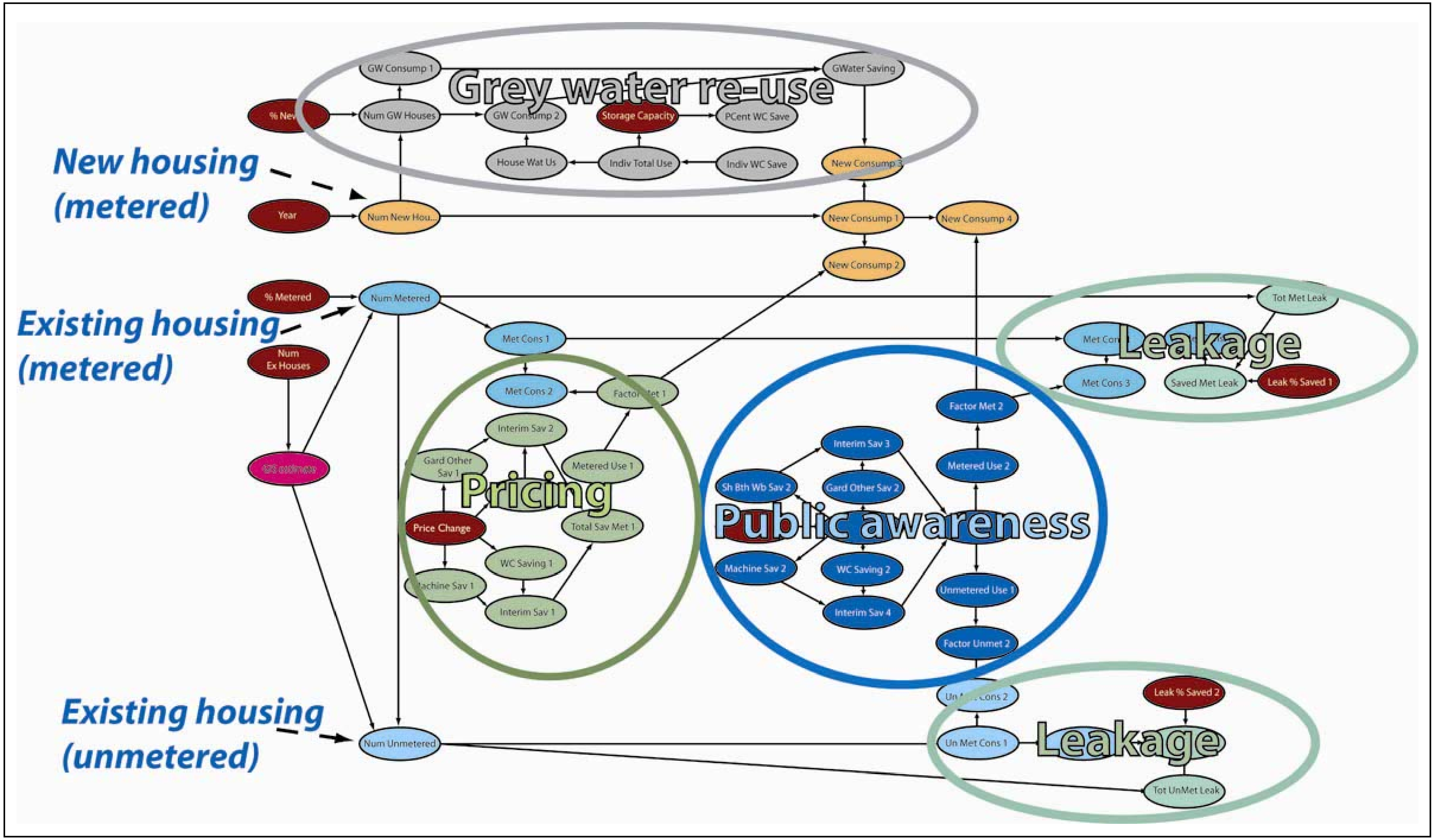


Figure 5.3 The demand management network for the Loddon catchment, highlighting the main types of action available to control demand

The network also examines the potential effect of introducing grey water systems in new housing stock. These systems store the waste water from sinks and baths to use in the toilet for flushing. With a standard storage tank of 20 litres the results show that a 26% reduction in consumption can be achieved. The savings offered through this strategy are considerable, but require legislation to ensure that new housing stock is fitted with the system, and of course there is a cost implication. There also needs to be education of the public to accept the use of waste water because experience has shown people are reluctant to use 'dirty' water in the house, even for flushing.

5.2 Compensation network: Denmark

Background

Water for the Copenhagen region is supplied by Copenhagen Energy (CE) from 55 well fields in North Zeeland. The area from which groundwater is drawn is one of the most densely populated and intensively farmed regions in the country. Extensive use of pesticides and fertilizers by local farmers has led to a serious contamination threat to the aquifers. In response to the threat CE are examining a number of options to safeguard supplies.

Among these options is the establishment of groundwater protection zones in the vicinity of well fields through the implementation of voluntary 5 -10 year farmers' contracts. These contracts, between the farmers and CE, guarantee no (or reduced) pesticide application in return for compensation payments. By restricting or banning the use of pesticides within an agreed distance of ground water sources, a zone is established within which groundwater is protected from pesticide pollution. Copenhagen Energy collaborated with MERIT to use Bayesian networks to (a) investigate the level of compensation payments required to create sufficiently large groundwater protection zones to be effective, (b) to identify the impacts of protection zones on land-use, farm economics, groundwater quality, biodiversity and the aquatic environment of the region, and (c) to develop a technique to encourage and enable full stakeholder involvement in the decision making process through the use of Bns. The stakeholder objective was seen to be particularly important in the light of potential conflict between CE and the farmers.

The study area was the Havelse well field, part of the Slangerup waterworks located in the north east of the North Zeeland district (Figure 5.4). The well field was installed in 1955/6 and is part of CE's investment plan from 2002 to 2006. The applicability of Bns to the problem of developing preventive groundwater protection plans, with a focus on pesticides, was investigated for the Havelse well field capture zone in North Zeeland. (5.4).

Stakeholder involvement

A list of water users, potential groundwater polluters, and water authorities was drawn up to help identify potential stakeholders in the area; these included waterworks companies, water consumers, farmers, industry, anglers and local governing bodies. All professional organisations with a potential, or even marginal, interest in groundwater protection in the area were invited by letter to a one-day workshop in October 2002. It is noteworthy that the more 'green' NGOs did not respond, while the 'Industrial' stakeholders indicated they preferred to use their political contacts on groundwater issues. Following the workshop a 'professional' stakeholder working group was formed, comprising 10 institutions including the project end-user CE, the local Agenda 21 Centre (facilitator and project co-ordinator), and GEUS (project owner). These institutions are listed in Table 5.2.

	Stakeholder	Rights	Roles	Responsibilities for water management
1	Copenhagen Energy	A municipality and waterworks (water supply company owned by Copenhagen Municipality)	Main interest in protecting groundwater resources from pollution. End user in MERIT project	Water supplier for Copenhagen Municipality and other areas in Copenhagen
2	NOLA, North Zealand Farmers Union	Participant in co-ordination forums and in co-operation forums, NGO	Advisory centre and local political branch of the larger farmers' association.	Agricultural sector manages use of more than 60% of Danish land. Agriculture linked to decentralised GW abstraction and protection
3	SJFL, Zealand Family Farmers' Association	Participant in co-ordination forums and in co-operation forums, NGO	Advisory centre and local political branch of the minor farmers' association	See above …
4	Frederiksborg County	Issue water abstraction licences. Decide water management plans	The water resources authority for both groundwater and surface water	Sustainable use of water resources; protection of groundwater resources; water resources plans; conduct public hearings
5	Frederiksværk Municipality	Issue water supply plans. Issue Groundwater protection action plans	Owner of local public waterworks	Initiate action plans for groundwater protection
6	Skævinge Municipality	See above …	See above …	See above …
7	FVD, Organisation of private waterworks in Denmark	Issue additional action groundwater protection plans. Participant in co-ordination forums and in cooperation forums, NGO	Political organisation for private waterworks. NGO	Initiate action plans for groundwater protection. NGO
8	Green Forum Slangerup	NGO	Associated to local Agenda 21 in Slangerup. Umbrella for local 'Green NGO's'	Committed to work for an alleviation of human pressures on environment
9	Agenda 21 Centre	Issue initiatives based on Agenda 21 principles	Agenda 21 Centre for 5 local municipalities	See above …
10	GEUS Geological Survey of Denmark and Greenland	Government organisation	Danish co-ordinator of MERIT	Scientific interest in water management and stakeholder involvement

Table 5.2 The Professional stakeholder group

A second stakeholder group of 'local citizens' was also formed. Invitations were sent to 1100 local households to attend a public meeting, which was also advertised in the local newspaper. The response was positive. Over 100 people attended along with the local television station. Those present were asked to identify the groundwater protection issues they considered to be important. At the end of the meeting, based on the responses received, a 'local citizen' working group of 9 persons was formed.

The split into a 'professional' and 'local citizens' group was made to ensure that professional stakeholders did not unduly influence the contributions of less experienced and knowledgeable non-professional individuals. A facilitator from the local joint municipality Agenda 21 Centre was used at all meetings.

Three workshops were held with the professional stakeholder group. The first was used to clarify roles and responsibilities and to elicit opinion about the different measures to actively protect groundwater. The next two workshops were used for Bn development, and included inputs from external experts. Following the workshops, individual meetings were arranged with Frederiksborg County Council and Sjællandske Familielandbrug to collect more data, and to discuss the Bns in more detail.

The citizen group met 5 times. All meetings were facilitated, but GEUS and CE only participated in 2 of the meetings. The group published 2 newsletters, which were distributed to 1000 households in the local area. The newsletters introduced members of the group and included articles on groundwater protection, water supply, and water quality. The citizen group will continue after the end of the MERIT project as a ‘water group’. In the final phase of the project several individual meetings were arranged to collect quantitative data for the Bns and complete the networks for farming contracts and flooding.

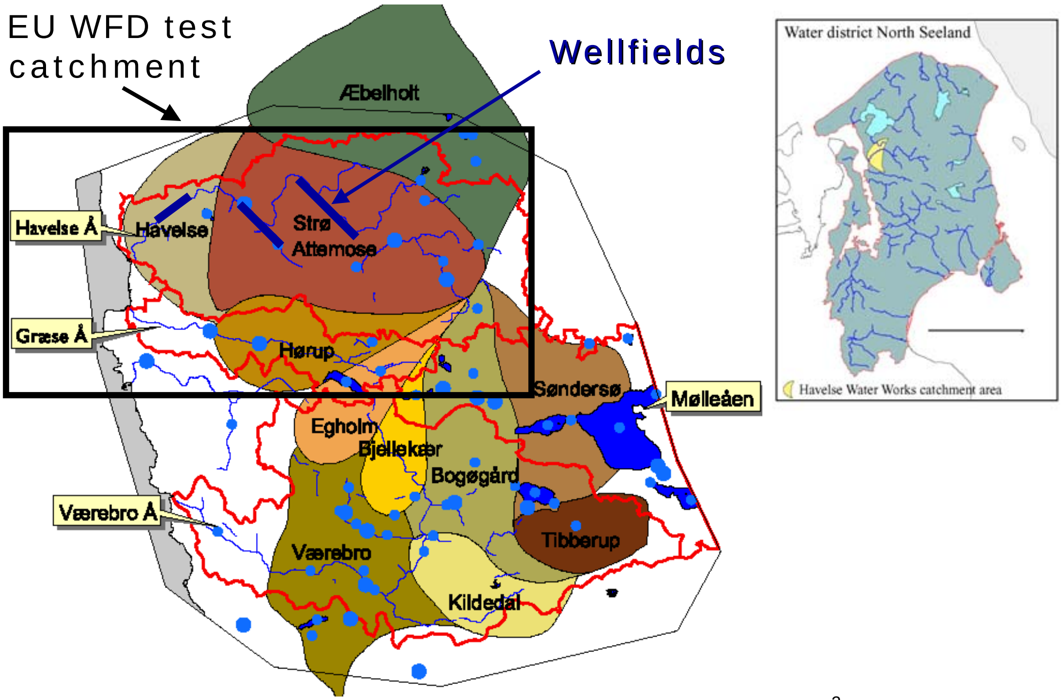


Figure 5.4 Danish case study: Havelse river catchment (~ 100 km^2). Havelse well field capture zone (35 km^2) located at the downstream end of the river basin near Roskilde fiord.

The compensation network

The Danish study included the development of 3 networks, but only the ‘compensation network’ is described here (Figure 5.5). The aim of this network was to analyse the effects of compensation payments to farmers for not using pesticides on agricultural fields[13]. The higher the compensation level, the more farmers will join the scheme. However, the impact of compensation will not be restricted to the reduced application of pesticides; there will be additional effects. For example, farmers joining the scheme will inevitably grow more of the types of crop not requiring the use of pesticides, so that crop rotation patterns will be altered. At the same time

[13] Henriksen, H.J. Rasmussen, P., Brandt, G., Bülow, D.v., Jørgensen, L.F., and Nyegaard, P.(2004b) Test of Bayesian belief network and stakeholder involvement. Groundwater management and protection at Havelse well field in Northern Zealand. EVK1-2000-00085 – MERIT (Danish case study). Danmark og Grønlands Geologiske Undersøgelse and Københavns Energi.

other sources of non-agricultural or off farm income may be pursued. In other words compensation payments will have an affect on the structure of farm economics throughout the region.

All the relationships for the economic section of the network were provided by the Royal Veterinarian University, who also collected the data for pesticide application for different crop rotations. The remaining part of the network contains variables relating to the environmental impacts of pesticide application. These variables were based on information from monitoring programs run by GEUS and Copenhagen Energy. It was assumed that wetter conditions increased the leaching of pesticides to shallow groundwater bodies. Research from the US shows that high concentrations of herbicides in surface water impacts the reproductive capability of leopard frogs (expressed by the variable 'biological abnormality').

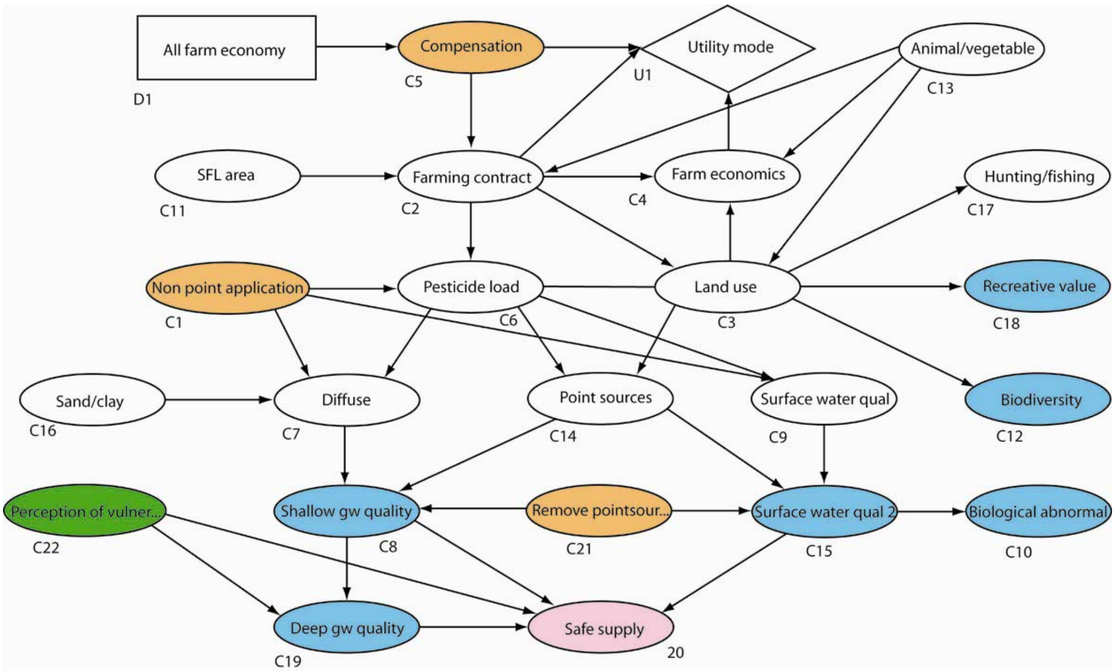


Figure 5.5 Danish 'compensation' network

One feature of the network worthy of note is the inclusion of a variable, '*perception of vulnerability*', that addresses the disagreement between farmers' organisations and the project leadership (GEUS and Copenhagen Energy) regarding the relationship between shallow and deep groundwater. The *'perception of vulnerability'* variable links to '*safe supply*', which describes the probability of having a supply of clean groundwater in the next fifty years. The result indicated in *'safe supply'* depends on which stakeholder point of view is selected in *'perception of vulnerability'*. The farmers' perception tends to be rather more optimistic than that of the water organisations. This is one way networks can be used overcome the stumbling block sometimes posed by conflicting points of view.

To help with the decision making process, a decision node 'All farm economy' (rectangular box) is included. This gives the total income (compensation) to the farmer, including subsidies, for different scenarios.

Results

In the network two main scenarios have been analysed:

1. The implementation of farming contracts: these are voluntary farming contracts at various levels of compensation
2. A joint action including voluntary farming contracts, plus the removal of point sources of pollution

The results of a series of network simulations shown in Figure 5.6 illustrate that for a safe groundwater supply to be guaranteed (95% certain), compensation payment has to be at least 4,400 Dkr. ha^{-1} yr^{-1}. This level of compensation can only realistically be paid under MVJ[14] agreements, which are granted under strict environmental rules. Reducing compensation dramatically reduces the number of farmers prepared to join the contract scheme; for example with compensation reduced to 2,500 Dkr ha^{-1} yr^{-1} recruitment levels fall to 50%, at 1,000 Dkr. to 11%, and at 500 Dkr. to a mere 4.1%.

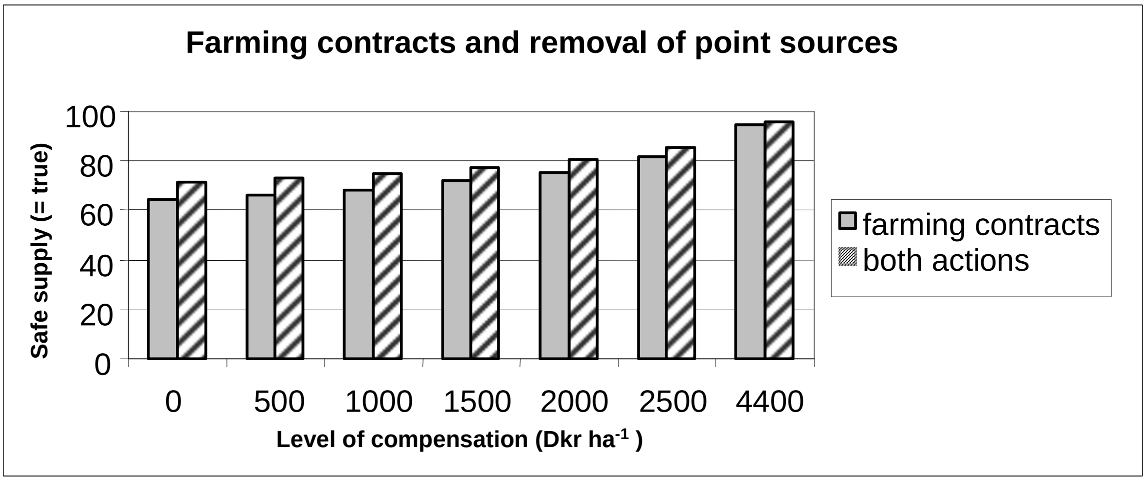


Figure 5.6: Comparison of overall indicators (safe supply) for two alternative scenarios: ‘farming contracts’ and ‘both actions’, under different levels of compensation

The network revealed the main problem to be the lack of commitment by the authorities to pay acceptable levels of compensation. Farmers in the area have suggested compensation payment levels of around 5,000 Dkr. $ha^{-1}yr^{-1}$. Farmers’ organisations have indicated that such agreements should either be set at a very high level (up to 7,000 Dkr. ha^{-1} yr^{-1}), or not form part of a groundwater protection scheme at all, suggesting that in practice expropriation may be a better method. The network results have convinced CE that compensation payments are not a realistic option for groundwater protection zone control; other possibilities are now being explored.

[14] MVJ is the Danish abbreviation for ‘Environmental friendly agricultural agreements’; these agreements demand zero pesticide application as well as imposing restrictions on the use of fertilizers and the choice of crop. The level of compensation for MVJ is DKr 4400 per ha/year, of this up to 60% is financed by the EU.

5.3 Irrigation network: Spain

Background

The 'Eastern La Mancha' aquifer in south east Spain supplies water to irrigate 105,000 ha, and meets the urban demand of 275,000 people (Figure 5.7). The total annual demand is 450 hm^3 (425 hm^3 for irrigation and 25 hm^3 for urban supply). Of this total, 70 hm^3 is supplied by surface water; the remaining 380 hm^3 is extracted from the aquifer.

The exploitation of groundwater has been the key to agricultural and economic development in La Mancha over the past forty years. However, since 1975 an expansion of irrigated areas has led to increased rates of abstraction and a dramatic decline in groundwater levels. As a consequence, abstraction from the aquifer is now unsustainable.

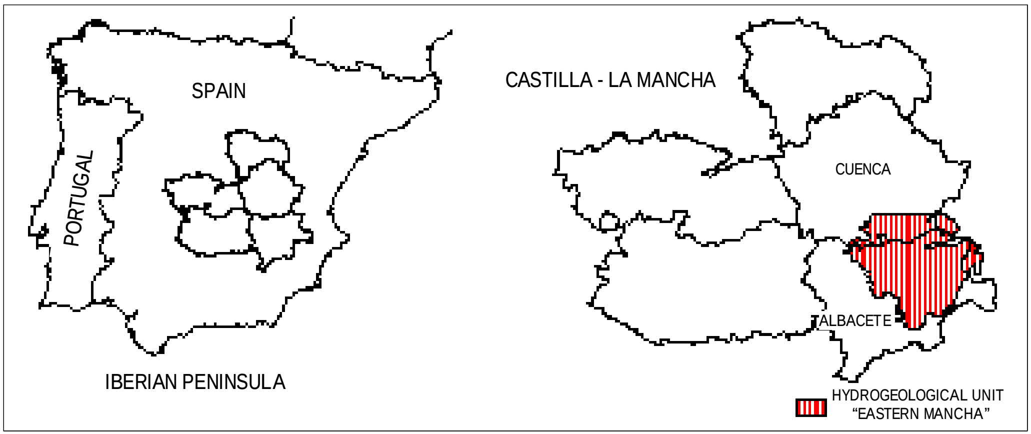


Figure 5.7 The 'Eastern Mancha' aquifer in Castilla-La Mancha, south east Spain

In response to this crisis an Irrigation Users' Association (JCRMO) was founded in 1994 by users of the aquifer. This organisation brings together farmers and other water users with a common aim: to achieve a sustainable level of water use in the aquifer unit. Irrigation accounts for more than 90% of total water consumption and is, therefore, the main target for improved management. Urban and industrial users account for less than 10%. But in addition to irrigation and urban use there is a third type of demand that needs to be considered; environmental demand. This is the water needed to maintain a healthy environment in the region, as well as that required for sport and recreation purposes. In this respect all citizens of the community have the right, as well as the duty, to express their views. Public representation on the JCRMO is made through the Ministry of Environment and various NGO organizations.

The aim of this case study was to construct a Bayesian network to help develop a sustainable strategy for the East La Mancha aquifer. This was achieved in partnership with the main managers and users of the aquifer.

Stakeholder involvement

Questionnaires and leaflets were sent by post to identify relevant stakeholder groups. Based on the responses, representatives were invited to attend a series of meetings and workshops. The selected representatives met a number of requirements: they lived or worked in the area, were able to participate in meetings and workshops, had some authority within the group they represented, and were well acquainted with the problems to be addressed. Once selected, an initial meeting was held to explain the aims of the project, the method to be adopted, and what was expected from those who might decide to participate. Participants were given a questionnaire to identify (a) their role as a user, regulator or expert (b) their decision-making capacity (c) any information they possess about the aquifer and (d) their opinions about the most appropriate way to improve water resources management in the aquifer.

The information from the questionnaires was used to construct a preliminary network. Managers and users of the aquifer were then invited to a series of meetings to examine and discuss the results. These meetings were attended by representatives of the Hydrographical Confederation of the River Jucar (CHJ), which is the public authority responsible for managing the aquifer resources, members of the JCRMO, and the Provincial Technical Institute of Albacete (ITAP). The latter is a body which specializes in the transfer of technology. Meetings were chaired by the members of the University of Birmingham, specialists in stakeholder involvement. The participation of these bodies in the development of the tool accounts for the high level of confidence in its results. The CHJ, JCRMO and the ITAP have all expressed interest in applying the technique to a number of other water resource problems in the region.

Network development

The network developed is shown in Figure 5.8. Selection of variables was based on the experience of each team member, and on the information from questionnaires completed by users and experts.

The network comprises 56 variables arranged into 8 groups, representing: (1) water inputs to the aquifer (2) environmental requirements (3) urban consumption (4) irrigation consumption (5) irrigation efficiency (6) maximum water available for irrigation (7) policy and legal tools, and (8) agrarian income. This configuration emphasises the importance of agriculture in the management of the resource. In comparison to agriculture the requirements of urban and environmental uses are small and can easily be met by current volumes entering the aquifer.

Results

The network operates by deducting river flow, urban demands, and agricultural requirements from the initial input volume. Although river flow (environmental) requirements are not consumptive they nevertheless limit the actual availability of water.

To be sustainable the total annual abstraction from the aquifer needs to be less than 320 hm^3 $year^{-1}$. Current levels are far in excess of this figure. The network was used to identify the impact of a number of different management scenarios. These scenarios involved changing the pattern of crop type and distribution, augmenting irrigation supply with various amounts of surface water, changing the efficiency of irrigation systems, and examining the impact of various types of subsidy paid to farmers. The challenge was to reduce the amount of annual abstraction to below 320 hm^3 $year^{-1}$, while at the same time maintaining an acceptable level of income for the farmer.

The result for each scenario was analysed, with all possible variations taken into account. The analysis was undertaken jointly by the research group and stakeholders. Inclusion of the stakeholders in this step was vital because ultimately they decide whether the results offered by the model, both partial and global, are acceptable or not. To achieve this goal several stakeholder meetings were organised to examine the network output. Once evaluated, the stakeholders could be certain that everyone involved accepted the tool and results. This process helped to increases users' confidence in the process.

The network approach used in La Mancha is one that meets the main aims of the European Water Framework Directive (WFD). Management of the resource is integrated, seeks to achieve sustainability of the aquifer, and both managers and users of the resource have been actively involved in the development of the network. The result is that decisions based on the approach are accepted by stakeholders. The structure of the network can serve as a reference point for future work geared

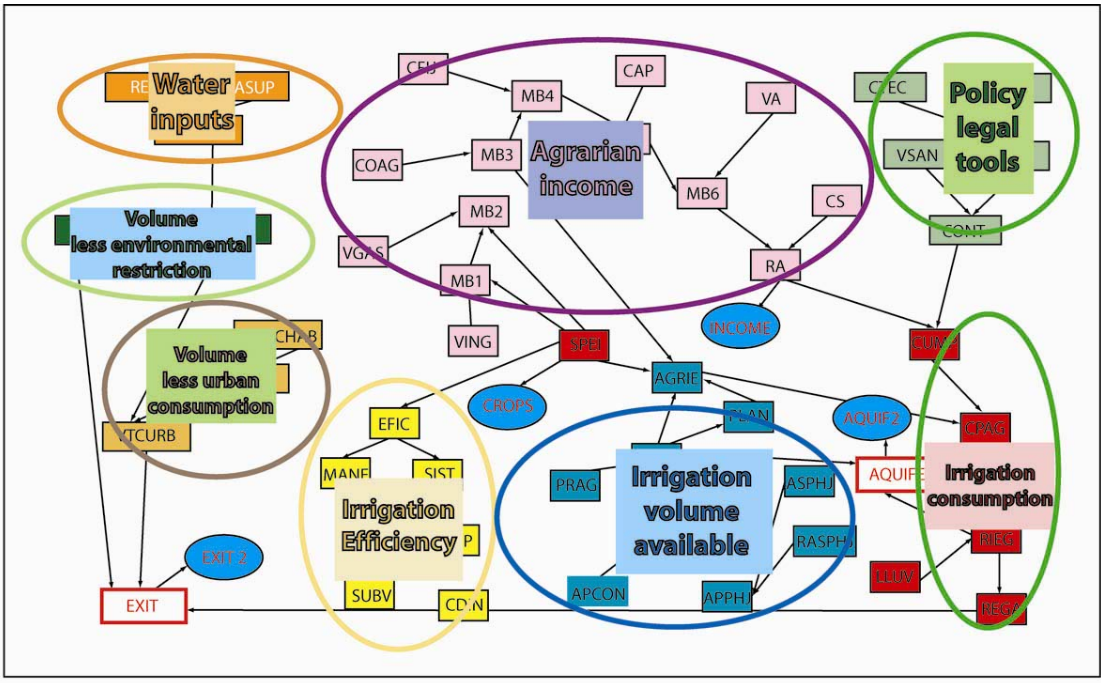


Figure 5.8 Bayesian network for the 'Eastern Mancha' Aquifer

to the integrated and sustainable management of ground or surface water resources, where the main consumer is agriculture. Proper use of the network, however, will require the users to be aware of its limitations and to display a critical attitude to the outcome.

5.4 Vomano network: Italy

The Vomano is one of four catchments in the Abruzzo region of central Italy. It covers 785 km² from the eastern slopes of the Apennines to the Adriatic Sea (Figure 5.9). Water in the Vomano is used mainly for hydropower generation and to a lesser extent for agricultural production and drinking water supply. Three dams form three major reservoirs along the course of the river (Figure 5.10).

Since the end of the Second World War ENEL, formerly the national power company (now privatised), has held all rights to water use in the Vomano. Naturally, the daily storage in the system is managed to meet the demands of hydroelectricity generation, which places it in direct conflict with other activities such as agriculture, environmental protection, and drinking water supply.

Figure 5.9 Abruzzo region and the boundaries of the Vomano water system (in red).

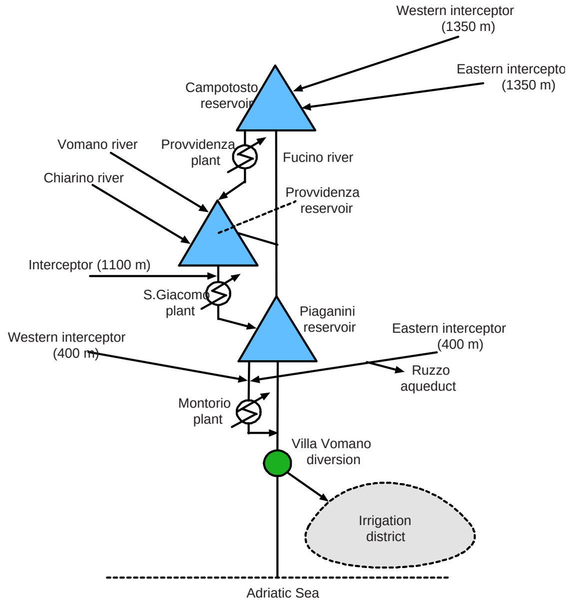


Figure 5.10 The Vomano system of reservoirs

Farmers and the hydropower company compete for available water at various times throughout the year. To meet increased demand during winter ENEL stores water during the late spring and early summer; but this is the time when crop growth is at its most rapid and when water is most needed by the farmers. It is also the time farmers need to fill the small reservoirs used for summer irrigation.

The reality is that the hydropower system frequently reduces flow in the Vomano to a trickle throughout the year, denying water to farmers and severely compromising the river ecosystem. For this reason environmental associations, as well as two national park authorities, are calling for the restoration of an acceptable level of environmental flow in the river.

These current conflicts could be made worse if farmers proceed with a plan to enlarge the irrigation district from the current 7000 ha to 14000 ha, and if the local government (Regione Abruzzo) decides to impose a minimum environmental flow in the river. As the decision maker, the local government is required to assess the impact of different management options on the hydrology, ecology, agriculture, water supply and hydropower generation in the catchment. The MERIT project has worked in collaboration with the Regione Abruzzo to use the construction of a participatory Bayesian network to address the problem.

Stakeholder involvement

The engagement of stakeholders in the evaluation process has taken place in three steps:

1. Elicitation – selection of stakeholders (mostly done through remote contacts)
2. Consultation – discussion of outcomes (meetings and telephone contacts)
3. Participation – selection of decision with the stakeholders (meetings).

The process began with a public meeting to identify (a) the physical and socio-economical boundaries of the system, (b) the existing and anticipated conflicts over water use and, (c) a set of 'indicators' to assess the impact of different management options.

Based on the response at the public meeting, a group of representative stakeholders were selected. The stakeholder group is shown in Table 5.3.

Stakeholder	Role
ENEL (Power Company)	Hydropower Production
Consorzio Bonifica Nord (farmers' league)	Agricultural Production
Acquedotto del Ruzzo (aqueduct)	Potable Water Supply
Parco Nazionale Gran Sasso – Monti della Laga	Environmental Quality Preservation
Regione Abruzzo (Water Authority)	Environmental Planning Preservation/DM
Provincia di Teramo	Environmental Quality Preservation
WWF, Legambiente	Environmental Quality Preservation

Table 5.3 The Vomano stakeholder group

A series of subsequent meetings of the group and sub-groups were held to identify scenarios, and to help design, contribute data to, and evaluate the network. A second public meeting was organised to present a provisional model of the system. Following this, a series of contacts with individual stakeholders, particularly the farmers' league and the power company, were made to help refine and improve the model to the point where it was acceptable to all stakeholders.

A final public meeting was held to present the results – i.e. the effects of each alternative on the set of indicators selected by the stakeholders. This stimulated a lively discussion between participants. Although it was not possible to have a real negotiation process due to the experimental aim of MERIT, the meeting was very useful for all the participants to widen their knowledge of the other components of the system. In particular, some conflicts that were originally assumed to derive from the structural features of the system were in fact demonstrated to be the effect of the management policies adopted in the river basin. Furthermore the meeting received a number of suggestions for improvements to the Bn model.

Network development

The network aimed to devise a water release strategy acceptable to both the electricity supplier and the farmers in the Vomano valley. The problem was to link the daily water transfers between reservoirs to the demands of farmers in a way that did not compromise electricity generation. Water transfers for electricity generation are easily calculated based on engineering principles. On the other hand, the water requirements of farmers is much more difficult to estimate and depends on a large number of factors, some of them related to human behaviour. To solve the problem, the reservoir operations model was coupled to a Bayesian network. Calculation of the daily water storages and movement within and between reservoirs was done using the reservoir operations model, whereas the more uncertain water demands of the farmers was assessed using a Bn. The combination of approaches proved to be extremely effective.

The representation of the irrigation district is shown in Figure 5.11. It contains four components:

1. Farmers' behaviour (Bayesian network)
2. Water distribution within the irrigated crops (release from the reservoirs)
3. The evapotranspiration process (Cropwat model)
4. Plant growth (FAO model)

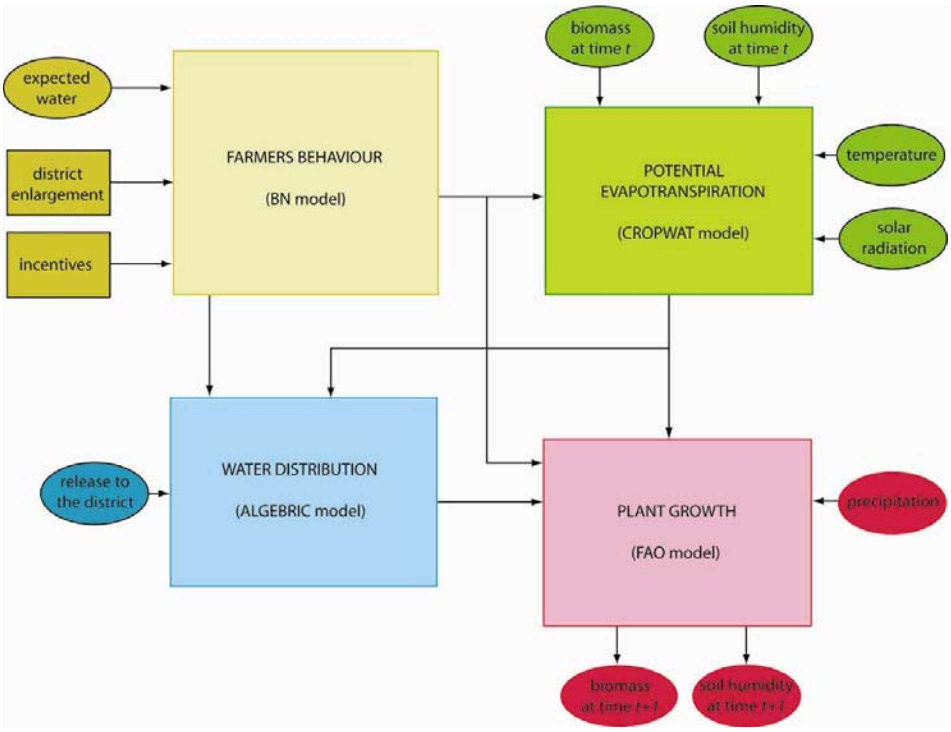


Figure 5.11 Components of the irrigation network.

The evapotranspiration, plant growth and water distribution components are derived either from models (Cropwat and FAO) or the measured releases from the reservoir. However, the component relating to farmer behaviour is obtained from a Bayesian network with inputs from the other three components (Figure 5.12). The variables in the behaviour component include financial incentives, the area of irrigation enlargement and the farmers' expectation in probabilistic terms that the water demand for agriculture will actually be satisfied, given the power company interests.

The outputs are obtained by fixing the value of the behaviour variables and running the model simulation for a given management policy. Different management policies release different amounts of water from the system over a given time period.

The CPTs were completed by a farmer league expert and during interviews with various farmers. The final Bn was validated by comparing the historical water uptake of the irrigation district against one generated by iteratively running the Bn (Figure 5.13).

Once designed and calibrated, the model of the irrigation district has been coupled to the operational model of the whole water system. Technically, the integration of the two models has been achieved by the implementation of an *ad hoc* programming code interfacing Hugin with a pre-existent C++ code through API Hugin libraries; in the process we lose the graphical facility of the Bn.

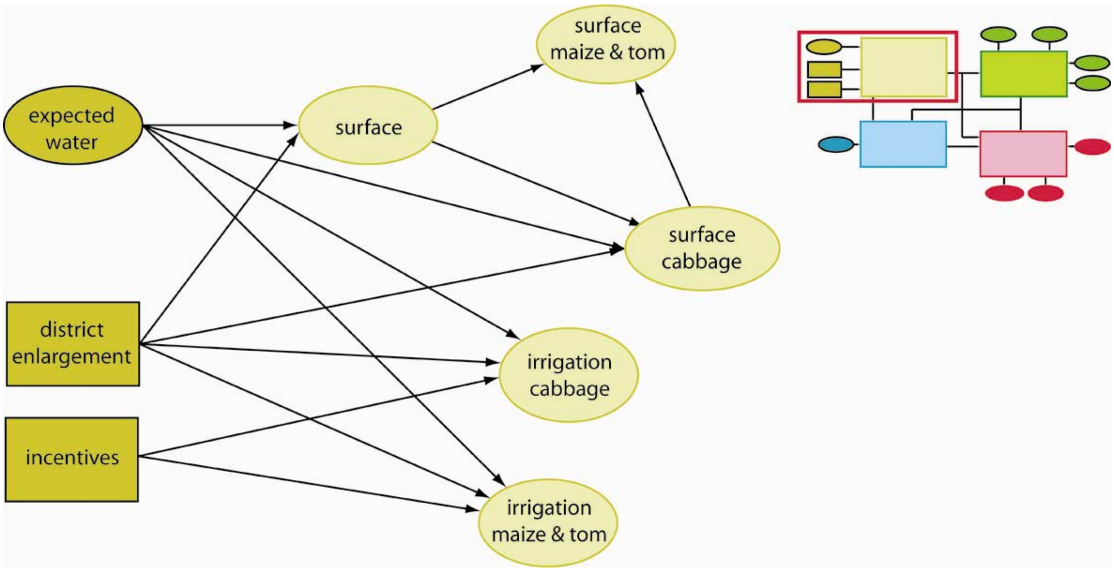


Figure 5.12 The Bn of farmers' behaviour

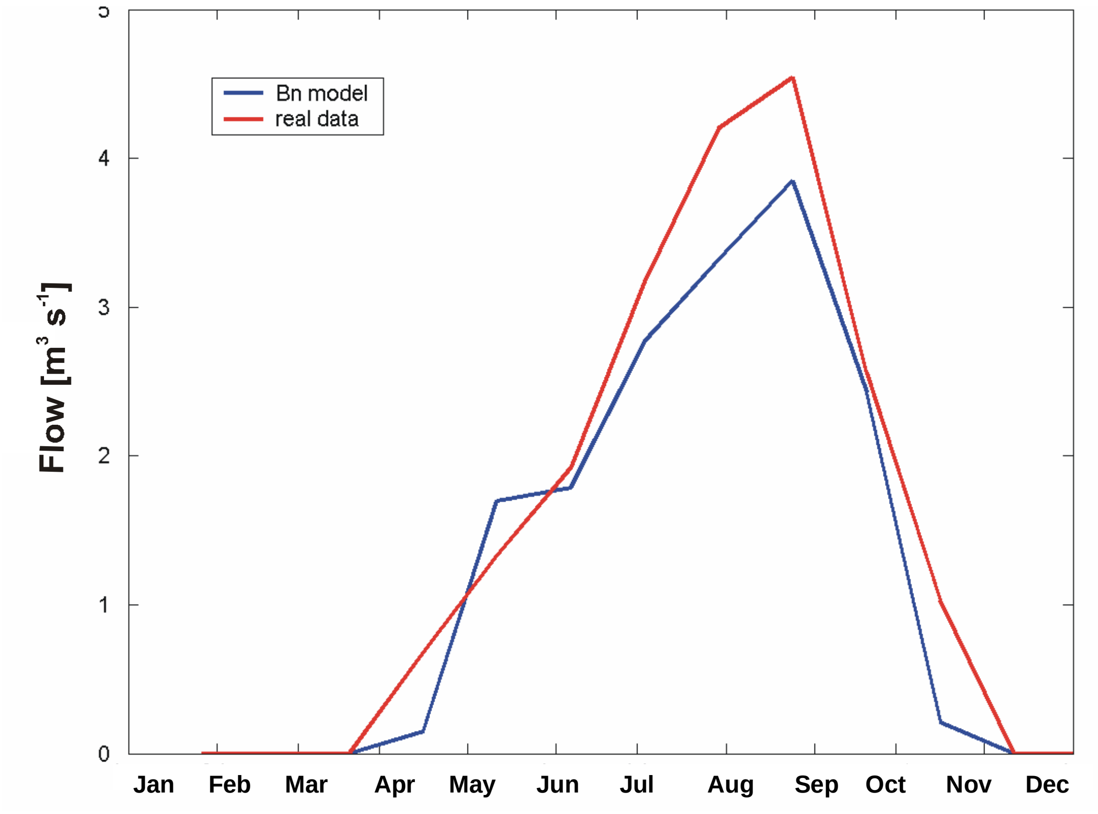


Figure 5.13: Validation of the Bn: actual versus the Bn generated uptake

A range of management scenarios have been simulated with the coupled system

1. Current situation, business as usual
2. Enlargement of the irrigation district
3. Imposition of a range of Minimum Environmental Flows (MEFs)
4. Doubling of the hydropower pumping capacity

5. Increasing of the Ruzzo aqueduct uptake to 1.4 and 2.1 $m^3 s^{-1}$.

The impacts of different scenarios on the hydropower company and farmers are shown in Figure 5.14.

IMPACT MATRIX	INDICATORS							
	Hydropower		Irrigation			Environment	Potable water	MEF
	Income [MIL Euro/y]	Production [Gwh/y]	Deficit [Mmc/y] on demand	Deficit [Mmc/y] on capacity	Ricavo S.Lucia [MIL Euro/y]	yearly lacking flow [mc/s]	Piaganini uptake [Mmc/y]	
ALTERNATIVES								
A0 History	34.77	527.65	16.16	12.35	0.43	0.50	3.74	1.00
A1 Current situation								
a. max. hydropower revenue	46.46	580.76	19.95	16.14	0.51	0.50	3.73	1.00
d. min agricultural loss lex to hydro	46.45	614.22	3.81	0.00	0.55	0.50	3.74	1.00
c. compromise policy	47.33	620.19	3.81	0.00	0.53	0.50	3.74	1.00
c.1 MEF interceptors								
c.1.1 interceptor 1100 (30%)	44.76	580.88	3.83	0.02	0.51	0.50	3.57	1.00
c.1.2 interceptor 400 (30%)	44.52	579.30	3.84	0.03	0.50	0.50	5.02	1.00
b. max hydropower rev. Lex agr.	47.43	605.02	11.89	8.08	0.51	0.50	3.73	1.00
b.1 Ruzzo uptake increase								
b.1.1 a 1.4 mc/sec	46.61	590.25	12.56	8.75	0.50	0.50	13.76	1.00
b.1.2 a 2.1 mc/sec	46.22	585.95	12.91	9.10	0.50	0.50	27.06	1.00
b.2 Fucino river MEF								
b.2.1 10% mean flow	47.29	602.07	11.85	8.04	0.51	0.45	3.73	0.90
b.2.3 30% mean flow	46.96	595.86	11.74	7.93	0.51	0.35	3.73	0.70
b.3 Mix of b.2.4 and b.1.2								
b.3.1 2.1 mc/sec + 30%	46.55	583.32	14.71	10.90	0.50	0.35	19.26	0.70
e. min agricultural loss	21.90	448.65	3.81	0.00	0.52	0.50	3.73	1.00
A2 2 X irrigation district								
a. max. hydropower revenue	46.46	580.76	44.84	44.84	1.04	0.50	3.73	1.00
d. min agricultural loss lex to hydro	40.97	592.94	0.03	0.03	0.97	0.50	3.66	1.00
c. compromise policy	47.55	644.38	0.96	0.96	1.01	0.50	3.72	1.00
b. max hydropower rev. Lex agr.	47.99	615.87	24.12	24.12	1.02	0.50	3.73	1.00
e. min agricultural loss	22.20	460.54	0.00	0.00	0.87	0.50	3.73	1.00
A3 2 X pumping								
a. max. hydropower revenue	48.19	602.20	25.46	21.65	0.49	0.47	1.10	1.00
d. min agricultural loss lex to hydro	46.49	607.64	3.81	0.00	0.53	0.50	3.42	1.00
c. compromise policy	47.17	604.44	3.81	0.00	0.50	0.47	3.42	1.00
b. max hydropower rev. Lex agr.	47.50	606.21	11.89	8.08	0.51	0.50	3.73	1.00
e. min agricultural loss	20.30	433.56	4.15	0.34	0.49	0.50	2.93	1.00

Figure 5.14 Impacts of a range of management scenarios. For each action, management policies optimising these objectives were considered:

A. Only maximum hydropower revenue (emphasis given to energy production)
B. Only minimum agricultural losses (emphasis put on agricultural production)
C. Minimum agricultural losses while maximising the hydropower revenue (combining the two previous objectives in a lexicographic sequence)
D. Maximum hydropower revenue while minimising agricultural losses
E. Both maximum hydropower revenue and minimum agricultural losses (compromise policy giving equal weight to energy and agriculture production)

The outcome for the current configuration of the system is presented in Figure 5.15. The Y axis shows the annual water deficit for irrigation in millions m^3, the X axis is the hydropower revenue in millions euros. The yellow circles represent the impact of the different alternatives shown in Figure 5.14; the red circle represents what has happened in the past. The results clearly show how a more efficient daily operation of the reservoir system could significantly increase water availability for the irrigation district, both in the current situation and in the eventuality of an enlargement of the irrigation district. For example, under current conditions option 1e and 1c provide a high yield for the power company, while minimising water deficits for irrigation. In contrast option 1a, while also maximising income, results in a significant irrigation water deficit. This result was surprising because it provides a good example of the way in which a conflict is caused more by a lack of efficiency in the current management system, rather than by an insurmountable obstacle.

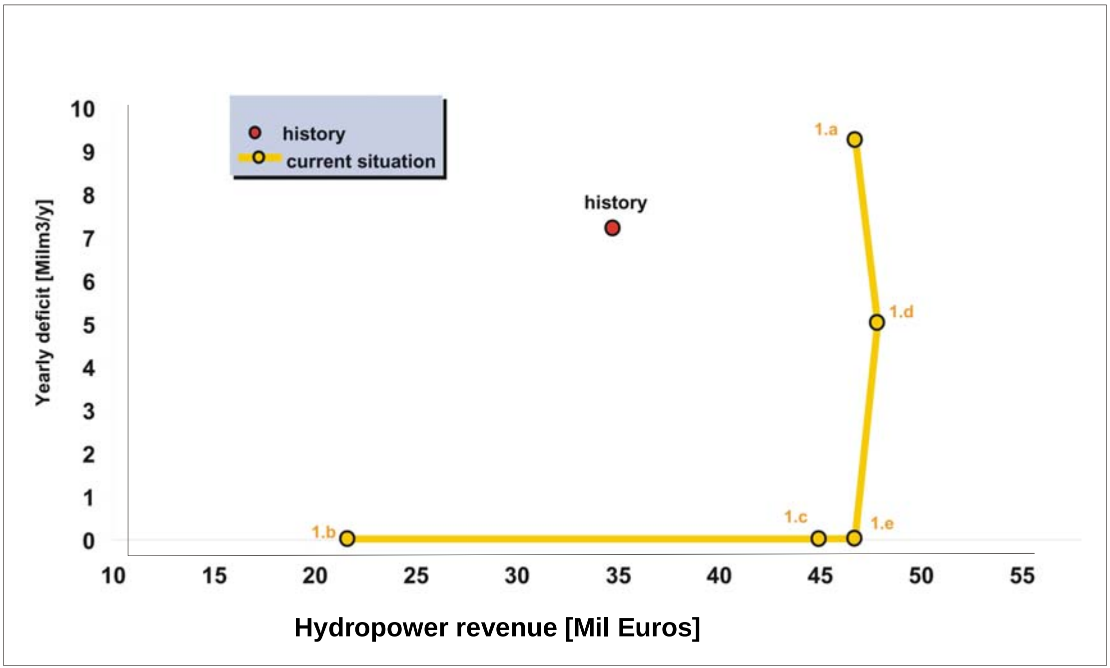


Figure 5.15 Outcome of alternatives for the current configuration of the Vomano system

Development of an environmental performance network

A second network relating to the Vomano was constructed to evaluate the environmental performance of different water management policies, with particular emphasis on the preservation of aquatic life. The network was developed for the Regione Abruzzo administration, which has the responsibility for the Regional Water Protection Plan. Stakeholder meetings were held to identify the main environmental problems caused by the management of the reservoirs. The stakeholders were consulted separately and included Local Authorities, National Parks, and the Anglers' Association.

The following issues were raised:

- River bank erosion caused by hydro-peaking (rapid flow variations due to daily adjustment of hydropower production)
- Problems connected to flow diversions in protected areas
- Water mixing across different watersheds

The conventional way to address environmental problems caused by water diversions is to impose a minimum release from dams (In-stream Flow Requirements – IFR). For this case study we have attempted to develop an index related to in-stream flow to help define the performance of various management options. The starting point is the microhabitat methodology for the estimation of IFR. A microhabitat simulation is a two-step process: firstly the microhabitat response to changes in flow is computed then secondly, the suitability of the new habitat is determined using a set of curves. The result is an index, called the Weighted Usable Area (WUA), with the dimension of an area. It represents an area weighted for fish preferences. It is not a measurable quantity but should be considered as an index.

The plot of WUA against discharge can be used to transform hydrological into biological information. The microhabitat methodology provides an estimate of the response of the aquatic ecosystem to different flows and is suitable for problems where there is a need to determine the effect of a perturbation on a specific river habitat. These characteristics make this approach suitable for the development of an environmental performance index for different water management policies.

Natural and regulated time series are taken into account by means of the respective duration curves. The curves of WUA versus flow allow the translation of the hydrological factors to a biological response. This allows a WUA duration curve to be derived starting with a flow duration curve which is usually easily available (Figure 5.16). The environmental cost of a water management policy is defined as an appropriate distance between the natural and regulated WUA duration curves. The performance index is called an ‘environmental cost’ because the higher the value, the worse the environmental impact would be. It is not meant to imply a monetary translation of the environmental impact.

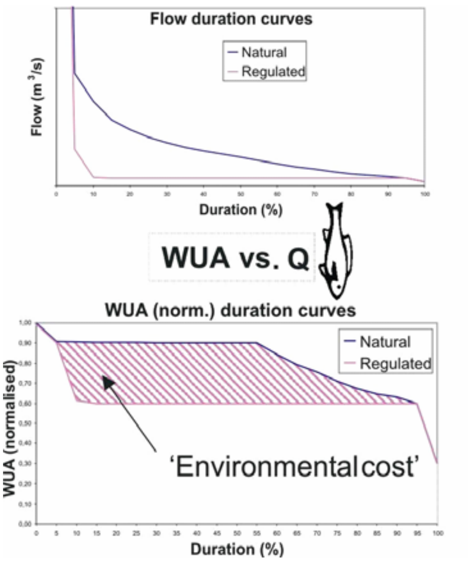


Figure 5.16 Derivation of a WUA duration curve and environmental cost from a flow duration curve

This methodology was used in a Bayesian network with the following objectives:

- Develop tool for the quantification of the ‘environmental cost’ of a management policy
- Develop a tool that allows stakeholders to assign their own weights
- Demonstrate use of the Bn approach in integrated water resource management

Figure 5.17 a-c illustrates the completed network. Green nodes are needed to transform area-based into point-based information. The points represent the locations from which water is released. The ‘biological species’ variable allows the user to select the fish of interest or to assign the relative importance of different species.

Pale green nodes transform a flow duration curve into a WUA duration curve, and can be used to select a subset of the WUA. Pink nodes allow switching between different policies and final indicators. Finally, the ‘stakeholder’ node can be used to select a predefined stakeholder.

Figure 5.17(a) shows the different ‘environmental costs’ of the two water policies from the point of view of the National Park; the cost indices are 10.69 and 10.33. Note the numbers in the ‘site’ node, representing the weights or importance given by the national park to various water derivation points, where minimum in stream flows will be released. Figure 5.17(b) shows the ‘environmental costs’ of the two water policies from the point of view of the Provincia di Teramo (a local authority). Apart from the different cost indices, it is interesting to note the numbers in the ‘site’ node, which are very different from the previous ones, indicating the derivation points are of different importance to each stakeholder. Finally Figure 5.17(c) displays the results for a generic stakeholder, interested only in the in stream flow at the Ponte Vomano

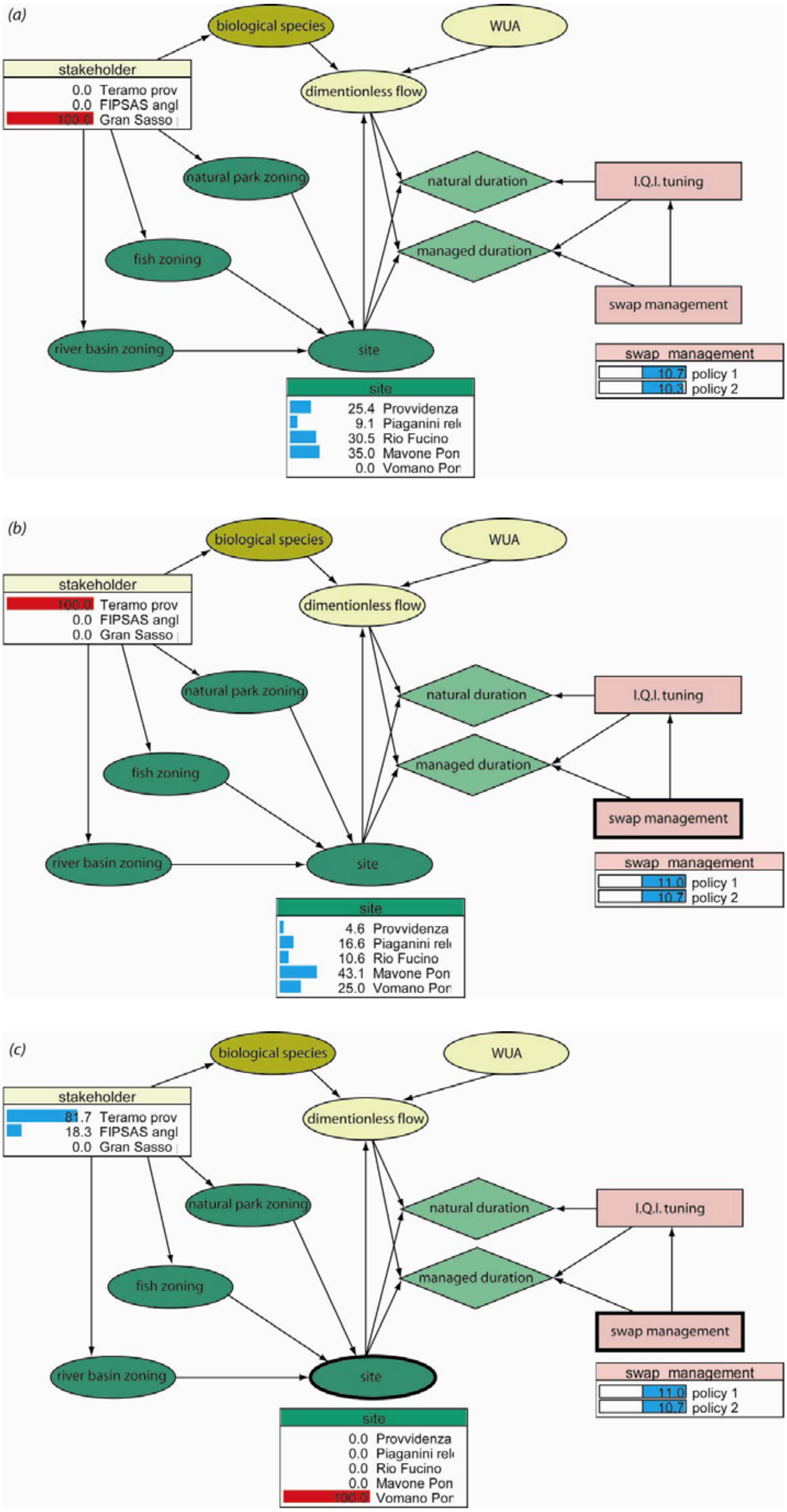


Figure 5.17 different ‘environmental costs’ of the two water policies from the point of view of (a) the national park; (b) the Provincia di Teramo; and (c) a generic stakeholder located at Ponte Vomano.

river section. In this case, the 'stakeholder' node can be read as an index that illustrates similarities between the generic stakeholder and the three predefined stakeholders. It is the Provincia di Teramo that would best represent the generic stakeholder's interests in this example.

6. BAYESIAN NETWORKS: strengths and weaknesses as a participatory water resource management tool

These guidelines describe the way in which Bayesian networks can be used to facilitate water resource decision making and improve the stakeholder engagement process. They have been based on the experience of four case studies. By and large the experiences were positive, though inevitably some difficulties along the way were encountered. Below we include a Strength, Weakness, Opportunity and Threat (SWOT) analysis of the Bn approach to water management, based mainly on the experiences of the MERIT project.

Strengths	Weaknesses	Opportunities	Threats
Bns are very visual; it is easy to demonstrate the way in which a system functions through the use of variables and links	Systems with a strong feedback element cannot be simulated	Opens up effective presentational opportunities, taking advantage of the visual nature of Bns. Provides a way to develop a clear understanding of the way system operates as a whole	Danger of being accepted without thorough understanding and appreciation of limitations
Can be used to integrate environmental, economic, social, cultural & political variables	Finding sufficient data to make the links between different disciplines – but this is a universal problem for all interdisciplinary approaches.	Provides a focus for bringing together different disciplines	Resistance of different disciplines to work together
Permits impacts of many different strategies and / or scenarios to be assessed in a short time	In large networks there is the danger of having too much information to take in.	A larger range of options can be evaluated	
Provides an excellent focus for dialogue with stakeholders	Danger of allowing the discussions to become unfocussed	New ideas and problems can be formulated quickly. Introduction of facilitator to direct stakeholder discussions	Discussions being dominated by a small number of articulate stakeholders
Good way to handle and interpret large data sets (structural and EM learning)	Difficult to understand for non experts		Open to manipulation
Can be used with incomplete data sets. Can also use a range of data including 'model output' and 'expert opinion'	Need to be aware of the basis of the cpts. If they are based on incomplete data uncertainties will be increased	Possible to evaluate systems lacking data	Placing too much reliance on uncertain data. Experts unwilling to give opinions
Good strategic and planning tool, but can also be used at smaller scales for specific problems	Not good at handling changes through multiple time steps	Problems can be addressed at a range of scales with a single tool	
Good predictive tool; the impact of a large range of future scenarios can be evaluated	Temptation to extrapolate beyond what is justified by the data		
Can be continually and rapidly updated; addition of new evidence reduces uncertainty	Need to have a well documented network to allow new operators to take over if necessary		Poorly documented networks can be difficult to interpret
Decisions made with a Bn are transparent when used in conjunction with stakeholder groups	Making sure stakeholders understand the principles on which Bns are based	Make environmental decision making more acceptable to the general public	Open to political manipulation

Glossary

Adder: A technique designed to reduce the number of links to a child so that the effects of different parents can be assessed independently.

Auxiliary variable: If a variable has large number of states it can be given a child auxiliary variable having only two states. This makes it easier to visualize the results of an action.

Bayesian network: A graphical tool used as an aid in decision making, particularly under conditions of uncertainty. Mathematically a Bayesian network is a graphical representation of the joint probability distribution for a set of discrete variables. The representation consists of a *directed acyclic graph* and to each variable *A* is attached the conditional probability of *A,* given the state of the parents of *A*. The joint probability distribution of all variables is then the product of all attached conditional probabilities.

Case: A piece of information that applies to one combination of the parent states.

Child variable: A variable that has links feeding to it from other (parent) variables.

Continuous variable: A variable that has a continuum of states. They are expressed as a Normal (Gaussian) distribution function described in terms of its mean and variance.

Controlling variables: Variables that cannot be controlled at the scale of the network, but have an impact on objectives e.g. rainfall, government policy.

Converging connection: Where a child has links from two parents.

CPT: Conditional Probability Table. A CPT for a specified variable provides a probability distribution over the states of that variable for each possible configuration of parent states (i.e. parent condition).

d-Connected: Where the path of evidence between two nodes is open.

Diagnostic mode: Where the state of a child is instantiated, and the evidence allowed to propagate back to the parent(s).

d-Separated: Where the path of evidence between nodes is blocked.

Discrete variable: A variable with a set number of defined states.

Diverging connection: Where a parent node links to two child nodes.

Divorcing: A technique to simplify a model by minimising the number of links made to a single variable through the introduction of a mediating variable or variables.

EPT: Elicited Probability Table. This is a CPT in which only a subset of the table is completed using input from experts or stakeholders; the remaining values are calculated using interpolation factors.

Indicator (objective) variable: This is a variable that you are attempting to control. It is also described as an objective variable and is used as a means to judge the success or otherwise of an intervention.

Influence diagram: A Bayesian network with utility functions and variables representing decisions.

INRM: Integrated Natural Resource Management.

Instantiate: To define the state of a variable; usually applies to implementation or controlling variables.

Intermediate variable: Variables that link interventions to objectives.

Intervention: An action that can be implemented to achieve an objective.

IWRM: Integrated Water Resource Management. An approach to water management which takes a 'holistic' view in which all relevant factors likely to be impacted by strategies are taken into account, and where the ultimate goal is sustainability.

Learning: The process of learning conditional probability distributions from an existing set of cases.

Link: Links connect the nodes of a network and indicates a causal dependence from the parent to the child node. The link is directed from cause to effect.

Modifying parent: A parent variable whose effect on its children, or some of them, is dependant on the states of other parents of those children.

Node: The element of a Bayesian network that represents a variable; the term variable and node are interchangeable.

Non-Modifying parent: A parent variable whose effect on its children is not dependant on the states of other parents of those children.

Objective variable: see indicator variable

Parent variable: A variable which links outward toward another (child) variable.

Serial connection: Where an action is transmitted through a series of variables

Stakeholder: A person or organization having a vested interest in the outcome of a decision either directly or indirectly; they can either affect or be affected by the decision.

State: All variables must have at least 2 qualitative or quantitative states. These are user defined and must cover all the states the variable can adopt, and they must not overlap. They can be expressed as a description, a value, a range or in Boolean form (e.g. True, False).

Scenario setting variable: A variable describing different a range of different scenarios.

Variable: See 'Node'

References

Bromley, J., Jackson, N.A., Clymer, O.J., Giacomello, A.M. and Jensen, F.V., 2004. The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning. Enviromental Modelling and Software, 20(2): 231-242.

Cain, J., 2001. Planning improvements in natural resources management. Centre for Ecology and Hydrology, Wallingford, UK, 124pp pp.

Huber, T., 1997. Introduction to Monte Carlo simulation. 1997

Jain, L.C., Takefuji, Y. and Johnson, R.P., 1998. Knowledge-based Intelligent Techniques in Industry. CRC Press, 352 pp.

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

Marakas, G.M., 1999. Decision Support Systems in the 21st Century. Prentice Hall, Upper Saddle River, New Jersey, 506 pp.

McLachlan, G.J. and Krishnan, T., 1996. The EM algorithm and extensions. Wiley Series in Probability and Statistics. Wiley-Interscience, 304 pp.

Neapolitan.R.E, 2003. Learning Bayesian Networks. Series in Artificial Intelligence. Prentice Hall, 674 pp.

O'Callaghan, J.R., 1995. NELUP : An introduction. Journal of Environmental Planning and Management, 38: 5-20.

Refsgaard, J.C. and Storm, B., 1995. MIKE SHE. In: V.P. Singh (Editor), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado.

Singh, V.P., 1995. Mathematical Models of Small Watershed Hydrology and Applications. Water Resource Publications, Highlands Ranch, Colorado, USA.

Turban, E. and Aronson, J.E., 2000. Decision Support System and Intelligent Systems. Prentice Hall, Upper Saddle River, New Jersey, 867pp pp.

UNDP, 1977. Report of the United Nations Water Conference; Mar del Plata, United Nations, New York, USA.

UNDP, 1992. Agenda 21, Chapter 18: protection of the quality and supply of freshwater resources; application of integrated approaches to the development, management and use of water resources., United Nations Conference on Environment and Development. UNDP, New York, USA, Rio de Janiero.

Varis, O., 1995. Belief networks for modelling and assessment of environmental change. *Environmetrics,* 6: 439-444.

Varis, O., 1998. A Belief network approach to optimisation and parameter estimation: Application to resource and environmental management. Artificial Intelligence, 101((1-2)): 135-163.

Vertessy, R., O'Loughlin, E., Beverly, E. and Butt, T., 1994. Australian experiences with the CSIRO Topog model in land and water resources management., Proceedings of UNESCO International Symposium on Water Resources Planning in a Changing World; June 28-30, 1994, Karlsruhe, Germany, pp. 135-144.

WMO, 1992. Development issues for the 21st century 26th - 31st January: The Dublin Statement and report of the conference, International Conference on Water and the Environment. WMO, Geneva, Switzerland, Hydrology and Water Resources Department, Dublin.

Appendix I

Basic statistical background of CPTs and properties of Bns

Basic statistical background

The concept used for the treatment of certainty in Bayesian networks is that of conditional probability. A conditional probability statement is of the following type:

…… if the variable B is in state b_1, then from either evidence or experience, we know that as a result, the probability of the variable A being in state a_1 is x. The notation for this statement is:

$$P(a_1|b_1) = x$$

The expression P(A|B) denotes a CPT containing numbers $P(a_i|b_j)$. Using Table A1.1 as an example, i = 1 to 3 and j = 1 to 3; in other words variables A and B both have 3 possible states. It should be noted that $P(a_1|b_1) = x$ means that whenever B is in state b_1, the probability of A being in state a_1 is x, *provided that everything else that is known is irrelevant for A.* This is important to remember, since other factors may have a significant affect on variable A. For example, in the case of Table A1.1 the variable 'Reservoir Storage' may also be affected by other factors such as 'Water Abstraction'. If this is the case, then this factor should also be built into the CPT. For our example, however, only the 2 variables 'Reservoir Storage' and 'River Flow' are considered.

		River Flow (Variable B)		
		Good (b_1)	Acceptable (b_2)	Bad (b_3)
Reservoir Storage (Variable A)	Good (a_1)	*0.9*	*0.6*	*0*
	Medium (a_2)	*0.1*	*0.3*	*0.1*
	Bad (a_3)	*0*	*0.1*	*0.9*

Table A1.1 Example CPT showing P(A|B); taken from Figure A1.1. Note that the sum of each column is 1

Table A1.1 is a CPT taken from the network shown in Figure A1.1 (Figure 2.2 in the guidelines). It shows the probability of A being in any particular state (a_1, a_2, a_3) given a state of B (b_1, b_2, b_3); it can be written as P(A|B). In the example, if 'River Flow' is in state b_2 (acceptable), the probability of 'Reservoir Storage' being in state a_1 (good) is 0.6 (60%), a_2 (medium) 0.3, and a_3 (bad) 0.1. Note that each *column* must add up to 1. In Table A1.1 the probabilities given, are based on the fact that the variable B has a 100% probability of being in state b_1, b_2 or b_3. But in reality we are unlikely to be certain of the state of river flow; there will always be some uncertainty. For instance, in the case shown in Figure A1.2 (Figure 2.3 in the guidelines) the variable 'River Flow' is not entirely in one state, but has the following probability distribution; 'good' 0.8 (80%), 'medium' 0.15 (15%), and 'bad' 0.05 (5%).

This probability distribution of B, written as P(B), together with the values given in the 'Reservoir Storage' CPT (Table A1.1), can be used to calculate the resulting probability distribution for reservoir storage, P(A). To obtain this distribution Bns use the *fundamental rule*, which can be written as;

$$P(A|B)P(B) = P(A,B)$$

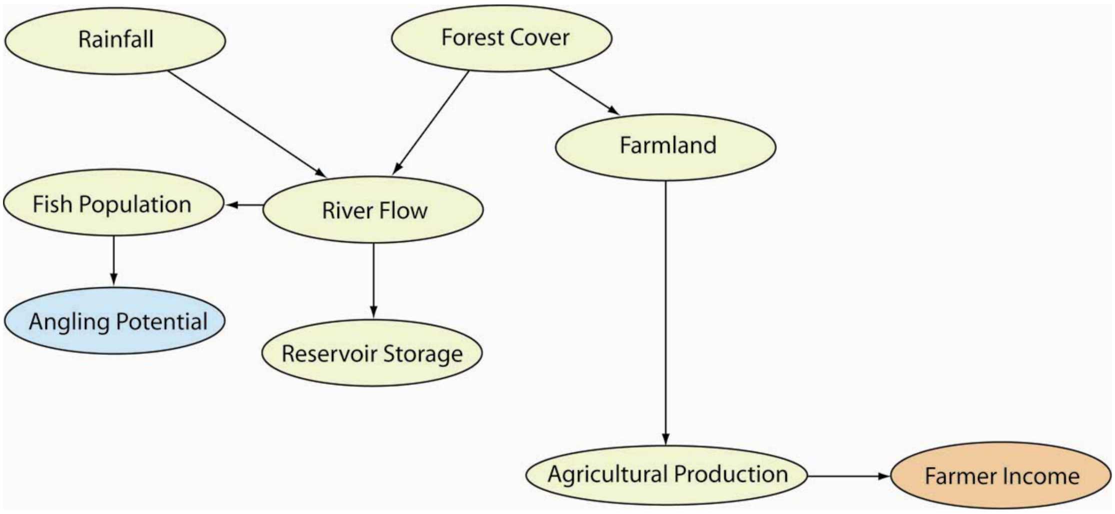


Figure A1.1 A simple Bayesian network

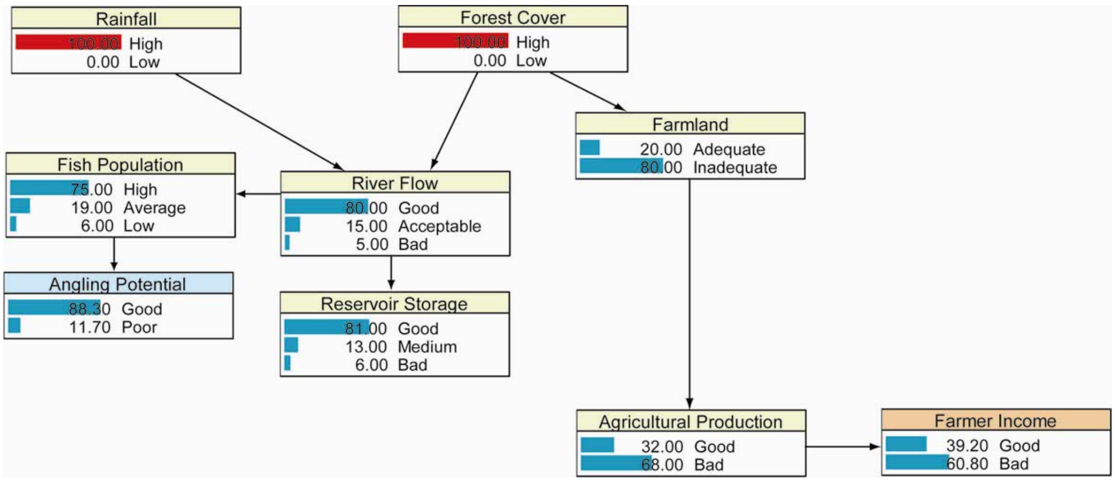


Figure A1.2 Compiled version of the simple Bayesian network shown in Figure A1.1

Here the term P(A,B) is an expression of the *joint* probability for the variables A and B. It consists of a table of all possible configurations (e.g. $a_{1\text{-}3}$, $b_{1\text{-}3}$ in Table A1.1). To construct the table P(A,B) it is necessary to know P(A|B), which we have from the CPT (Table A1.1), and P(B) the probability of B, given by the 'River Flow' node. Using the probability values for B (b_1=0.8; b_2=0.15 and b_3=.05), and the values in the CPT (Table A1.1) we can use the fundamental equation to calculate (P,A). Thus for the configuration a_1,b_1 we can write:

$$P(a_1|b_1)\ P(b_1) = P(a_1,b_1) = 0.9 \times 0.8 = 0.72$$

Values for all the other configurations can be worked out in the same way to produce the full P(A,B) table (Table A1.2). Note that in this table the sum of *all* the entries is 1. This Table gives the joint probability distribution for A and B, but what we really need is the probability distribution for variable A, Reservoir Storage'.

		River Flow (Variable B)		
		Good (b_1)	Acceptable (b_2)	Bad (b_3)
Reservoir Storage (Variable A)	Good (a_1)	0.72	0.09	0
	Medium (a_2)	0.08	0.045	0.005
	Bad (a_3)	0	0.15	0.045

Table A1.2 The *joint* probability distribution P(A,B). Note that the sum of *all* entries is 1

This distribution is derived using a calculation called *marginalisation* whereby the variable is B is marginalised out of P(A,B), resulting in P(A). From Table A1.2 it is clear that for any particular state of A (a_1, a_2 or a_3) there are 3 possible states of B (b_1, b_2 and b_3). Thus for each state of A there are 3 mutually exclusive conditions (a_1, b_1) (a_1, b_2) and (a_1, b_3). From this it follows that:

$$P(a_1) = \sum_{j=1}^{3} P(a_1, b_{1-3})$$

By marginalising B out of Table A1.2 we get:

$P(a_1) = 0.72 + 0.09 + 0 = \underline{0.81}$ (81%)
$P(a_2) = 0.08 + 0.045 + 0.005 = \underline{0.13}$ (13%)
$P(a_3) = 0 + 0.015 + 0.045 = \underline{0.06}$ (6%)

Examination of the compiled network in Figure A1.2 shows that these are the probabilities that appear in the 'Reservoir Storage' window for P(A).

Finally, it should be noted that the fundamental rule gives rise to the well known *Bayes' rule* in the following way:

$P(A|B)P(B) = P(A,B)$ ……………………….. *fundamental rule*

Thus it follows: $P(A|B)P(B) = P(B|A)P(A)$

Then: $$P(B\,|\,A) = \frac{P(A\,|\,B)P(B)}{P(A)}$$ ……………………….. *Bayes' rule*

Bayes' rule can be used to obtain the Table P(B|A), which is the CPT showing the likely state of the river flow given the reservoir storage; in other words the river flow is conditional upon reservoir storage, the reverse of situation shown in Table A1.1. Applying Bayes' rule to Table A1.1 gives Table A1.3. Thus in the case of $P(b_1|a_1)$:

$P(a_1|b_1)\ P(b_1)\ /\ P(a_1) = P(b_1|a_1)$

$0.9 \times 0.8 / 0.81 = 0.89$

The value $P(a_1|b_1) = 0.9$ is from the CPT in Table A1.1; $P(b_1) = 0.8$, the probability of river flow is given by the probability distribution of the 'River Flow' node shown in Figure A1.2, and $P(a_1) = 0.81$ is the probability given by the 'Reservoir Storage' node in Figure A1.2.

		Reservoir storage (Variable A)		
		Good (a_1)	Medium (a_2)	Bad (a_3)
River Flow (Variable B)	Good (b_1)	0.89	0.615	0
	Acceptable (b_2)	0.11	0.346	0.25
	Bad (b_3)	0	0.039	0.75

Table A1.3 CPT of P(B|A); obtained by applying Bayes' rule to the CPT of P(A|B) in Table A1.1. Note each column adds up to 1

Calculation of these equations is performed automatically by the Bayesian network software, so you will be relieved to know that it is not necessary to solve any of them manually. Nonetheless, although detailed knowledge of the statistical techniques is not essential to run networks, it is clearly preferable to have at least some understanding of the principles upon which the technique is based.

Some basic properties of networks

Direction and feedback loops

Although Bns are flexible and can be used to represent even the most complex systems, there are two basic properties that need to be observed during construction. Firstly they must be *directed*; that is they must act in one direction only. A connection between nodes A and B must be in the direction A → B, or B → A, but not in both directions at once. This is a basic property of networks that cannot be violated.

Secondly the network must also be *acyclic*. This simply means that links must not be allowed to form a closed loop or feedback cycle, because no efficient calculus has yet been developed to model this type of configuration. All networks must, therefore, be both directed and acyclic.

Conditional Independence and d-separation

The links between variables show the direction of cause and effect, but the way in which evidence is transmitted depends on the configuration of the connections, and whether or not evidence has been entered into key nodes. The type of configuration and the evidence available determines whether variables are dependent or independent, and it is important to be able to recognize the different types when they occur. When two variables are independent (i.e. the evidence about one does not affect our belief about the state of the other) they are said to be dependent – separated, or *d-separated* for short. When they are dependent (i.e. our knowledge of the state of one variable will affect our belief about the state of the other), they are described as dependent – connected, or *d-connected*.

In the following paragraphs we look at 3 types of connection - diverging, serial and converging connections - and examine the way in which evidence is transmitted through them.

Diverging connection

The situation shown in Figure A1.3 is a *diverging connection*. Influence can pass between the two child nodes, unless the state of the parent variable is known. This means that once we know the state of the parent (C in Figure A1.3) then any further evidence about B cannot change our belief about A and vice versa.

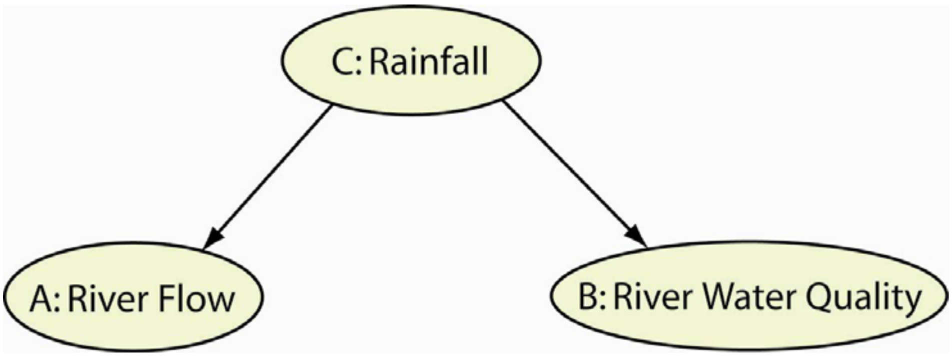


Figure A1.3 Diverging connection

In the example (Figure A1.3) the variables 'River Flow' (A) and 'River Water Quality' (B), are both dependent on 'Rainfall' (C). When rainfall is high, we know that river flow is also likely to be high, and water quality is probably going to be good because of the dilution effect of direct run-off. But suppose nothing is known about the state of rainfall, then evidence about either flow or quality will have an influence on the other. For example, if the quality is poor this might lead us to believe that it is more likely that flow is low and vice-versa; in other words A and B are *dependant* or d-connected.

However, if we have hard evidence that rainfall is 'high', then any evidence about water quality that we might have will not change our belief about river flow, which will be overwritten by our knowledge of rainfall. In this case 'River Flow' and 'River Water Quality' are said to be *conditionally independent* or *d-separated,* given the state of 'Rainfall'.

Thus evidence may be transmitted through a diverging connection, *unless the connecting (parent) variable is instantiated*

Serial connection

There may be a situation where variable A connects to B which in turn connects to C, a so called serial connection. A will have an impact on B, which in turn will impact on C. Clearly evidence about A will influence the certainty of B and in turn the certainty of C. In the same way evidence about C will influence our belief about B and A. However, if the state of B is known, then the route between A and C is blocked, in which case A and C are d-separated.

An example is given in Figure A1.4 where the variable 'Water Supply' (A) has an impact on 'Agricultural Production' (B) which in turn has an impact on 'Farmers Income' (C). Suppose all the variables have states 'high' and 'low', and we have some evidence that 'Water Supply' is 'low'. Then clearly this increases our belief that both 'Agricultural Production' and 'Farmers Income' will also be low. Evidence about A is thus transmitted through B to C. However, suppose that 'Agricultural Production' (B) is known to be in the 'high' state. In this case any information about 'Water Supply' (A) cannot be transmitted to the variable 'Farmers Income' (C) because it will be overwritten and replaced by the certain knowledge we have of B.

Thus in a serial connection evidence cannot be transmitted through the connecting variable if the state of the connecting variable is known

Figure A1.4 Serial connection

Converging connection

In the case where the links between variables A and B converge to C, then if nothing is known about C, the parents A and B are independent. But if the state of C is known, or even if we only have some vague indication about the state of C, then it is possible our beliefs about the state of A given B will be changed. An example of this type of converging connection is shown in Figure A1.5.

In this example, if the state of 'Recharge' is known, but there is no information about 'Water Resources', then nothing can be deduced about the state of 'Water Abstraction'; the variables A and B, are thus independent. If however, there is some evidence about the state of 'Water Resources', then from knowledge of 'Recharge' some inference about 'Water Abstraction' can be made. For example, if we knew that 'Recharge' was 'high' but 'Water Resources' was 'low', the implication is that 'Water Abstraction' is more likely to be in a 'high' rather than a 'low' state (to account for the low water resource). In this case A and B are dependent. Likewise if the state of D, 'Agricultural Production', a descendent of C, was known to be low, the same deduction would apply. It follows that in a converging connection evidence can only be transmitted between the parents A and B when the converging variable C, or one of its descendents, has received some evidence.

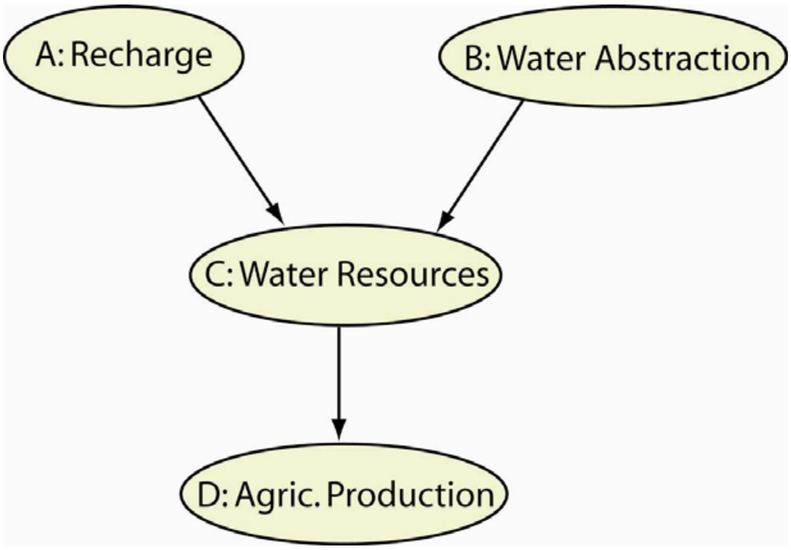


Figure A1.5 Converging connection

D-separation and d-connection: summary

In summary, two nodes A and B in a Bn are *d-separated* (independent) if, for all paths between A and B, there is an intermediate node C for which either:

1. the connection is serial or diverging and the state of C is known for certain; or

2. the connection is diverging and neither C (nor any of its descendants) have received any evidence.

If none of these conditions apply the nodes A and B are *d-connected* (dependant).

Appendix II

Completion of Conditional Probability Tables using large data sets

When you have a large data set the task of completing the resulting CPT for each case is daunting, particularly if there are a large number of states per variable.

Fortunately, commercial Bn packages, such as HUGIN, provide an automated process to calculate the probability for each case, based on the data available, and enters these into the resulting CPT. The procedure in HUGIN is known as EM (Expectation – Maximisation) learning and can best be illustrated by describing an example.

Example of EM learning

This example is a network that predicts the change in domestic water demand in response to changing weather conditions. In the summer months in the UK, Water Companies are aware that domestic demand is dependent to a large extent on (a) the temperature and (b) the rainfall. Water consumption will tend to increase during warm dry periods, but be reduced when the weather turns colder and wetter. The network provides a prediction of the likely consumption given different combinations of temperature and rainfall based on a 7 year data set of rainfall, temperature and water demand.

The network, shown in Figure A2.1, has 3 variables:

1. Average Maximum Daily Temperature in ^{o} C (AvMaxTdegC)
2. Weekly Rainfall in MM (WeeklyRainMM)
3. Domestic Water Consumption in ML day^{-1} (MLDay)

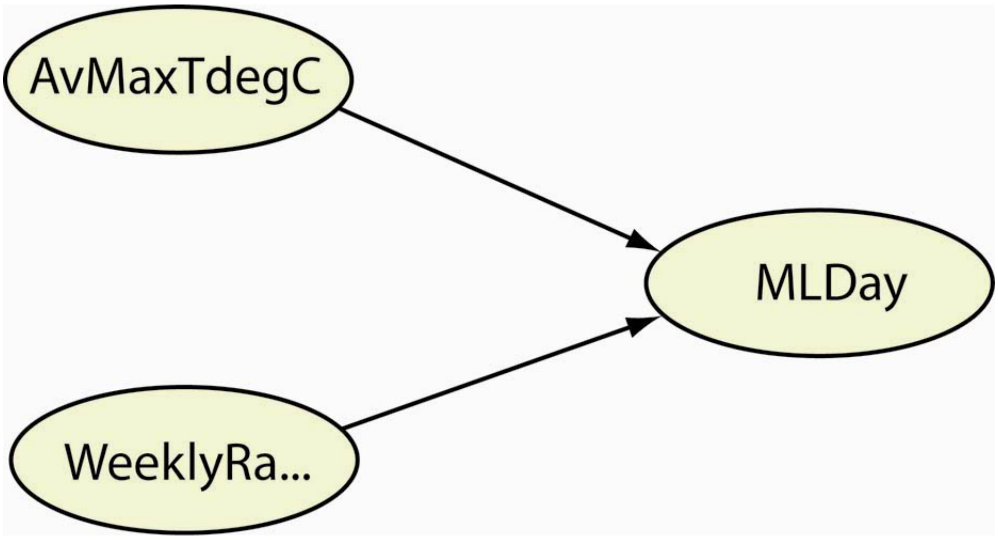


Figure A2.1 Domestic consumption network

The available data set has weekly rainfall, temperature and demand values covering a period of about 7 years. However, not all of this data is useful. In the winter months from October to March, the weather tends to be cool and wet and does not influence demand. For the construction of the CPT table, therefore, only data for the period from May to September is used. After making this adjustment the data set comprised 182 sets of information. The first 18 records are shown in Figure A2.2. The first column shows the average maximum weekly temperature, the second the weekly rainfall total, and the third the average weekly water consumption in mega-litres per day.

AvMaxTdegC	WeeklyRainMM	$MLday^{-1}$
10.21	41.1	211.30
10.22	22.7	212.16
13.92	0	220.65
12.28	9.5	220.47
13.56	0	227.66
15.03	0.6	229.80
14.20	4	220.86
14.36	25.9	225.53
15.54	9.1	213.13
21.67	0.6	243.43
16.47	33	218.07
16.91	0.4	230.79
20.64	0	262.21
20.62	0	273.82
16.80	20.8	225.12
18.00	13.5	225.95
18.05	16.6	222.25
18.34	1.2	230.86

Figure A2.2 Part of the consumption network data set

Once the structure of the network is defined and the relevant data obtained, the next step is to specify the states for each variable. In this instance the states are expressed as a series of intervals as shown in Figure A2.3.

Variable states

AvMaxTdegC	**WeeklyRainMM**	**$MLday^{-1}$**
0-15	0-5	210-220
15-20	5-10	220-225
20-30	10-15	225-230
	15-20	230-235
	20-30	235-240
	30-100	240-250
		250-300

Figure A2.3 The states for each variable; this gives a CPT with 126 cases (3*6*7)

With this combination of states, the CPT for the variable ‘MLDay’ requires the input of probabilities for 126 cases, clearly something that is impractical to do manually. Instead, the learning routine in the Hugin package can be used to generate and complete the table automatically using the data available. The procedure is as follows:

Step 1: The first step is to assemble the data in a tabular format that can be read by the software (e.g. 'txt' or 'csv'), similar to that shown in Figure A2.2.

Step 2: Initiate the learning wizard in Hugin. The first screen requests the name of the file containing the data.

Step 3: The next screen (Figure A2.4) is for data pre-processing. Here you are given the opportunity to select the variables you wish to include in the data set; in this case all three are selected. At this stage it is necessary to define the ranges of the states for each variable. This is done by highlighting the discretise option and entering the intervals for each variable in turn. In Figure A2.4 the intervals for the 'AvMaxTdegC' variable are shown.

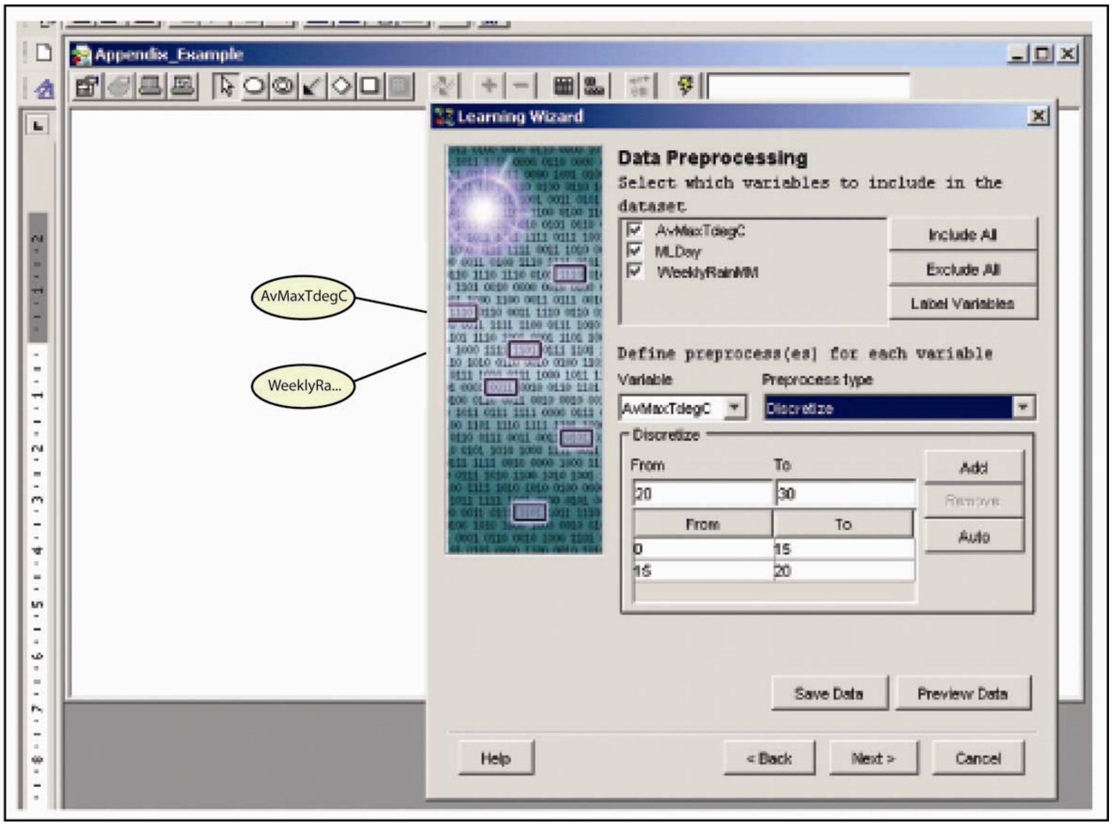
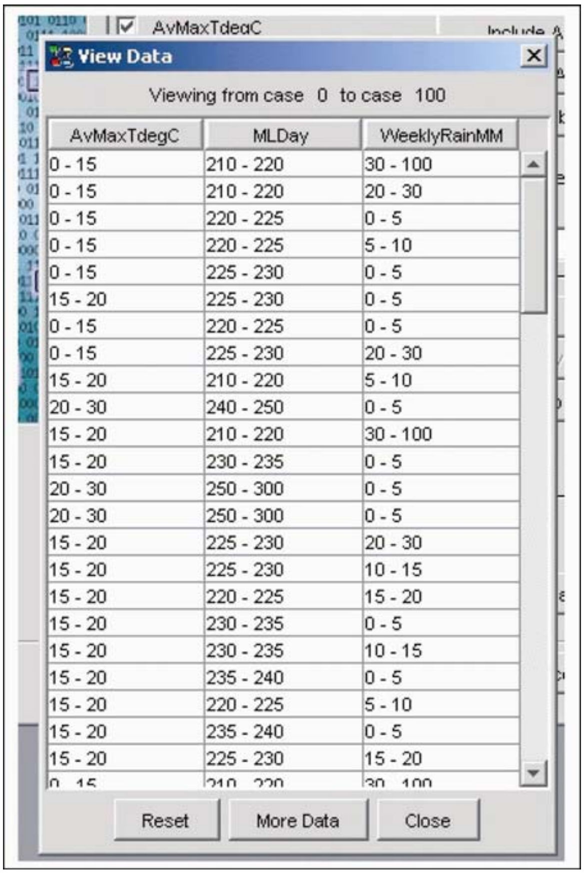


Figure A2.4 Pre-processing screen of the EM learning wizard in Hugin

Step 4: Once the states have been entered it is possible to view the resulting table, which now lists all the possible combinations of cases. Part of this file is shown in Figure A2.5.

Step 5: This step gives you the opportunity to specify the links between each variable. In this case we are sure about the links and have indicated where they should be placed. However, in some cases where you are not certain how the data links together, the software will suggest the strongest relationships based on the data provided.

Step 6: Here you are allowed to select the algorithm to be used to learn the structure of the data. The usual choice is the Necessary Path Condition (NPC) procedure; the level of significance (the probability of rejecting a true independence hypothesis) is normally set at .05 (Figure A2.6).



AvMaxTdegC	MLDay	WeeklyRainMM
0 - 15	210 - 220	30 - 100
0 - 15	210 - 220	20 - 30
0 - 15	220 - 225	0 - 5
0 - 15	220 - 225	5 - 10
0 - 15	225 - 230	0 - 5
15 - 20	225 - 230	0 - 5
0 - 15	220 - 225	0 - 5
0 - 15	225 - 230	20 - 30
15 - 20	210 - 220	5 - 10
20 - 30	240 - 250	0 - 5
15 - 20	210 - 220	30 - 100
15 - 20	230 - 235	0 - 5
20 - 30	250 - 300	0 - 5
20 - 30	250 - 300	0 - 5
15 - 20	225 - 230	20 - 30
15 - 20	225 - 230	10 - 15
15 - 20	220 - 225	15 - 20
15 - 20	230 - 235	0 - 5
15 - 20	230 - 235	10 - 15
15 - 20	235 - 240	0 - 5
15 - 20	220 - 225	5 - 10
15 - 20	235 - 240	0 - 5
15 - 20	225 - 230	15 - 20

Figure A2.5 Some of the cases defined by the EM learning Wizard

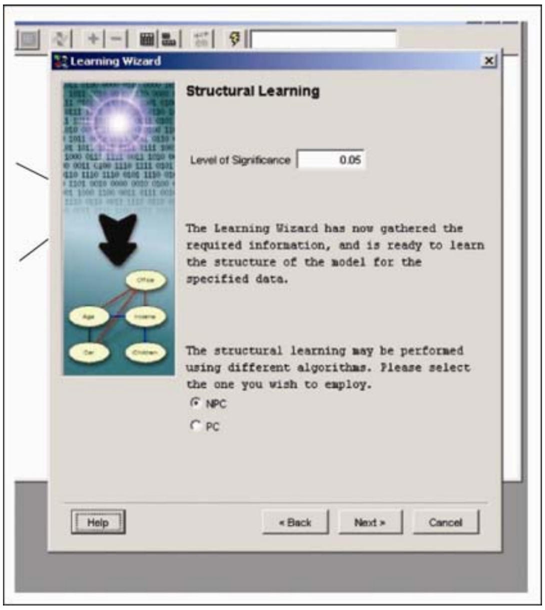


Figure A2.6 Selection of algorithm and significance level

Step 7: The next screen uses a slider to indicate the relative strength of the links.

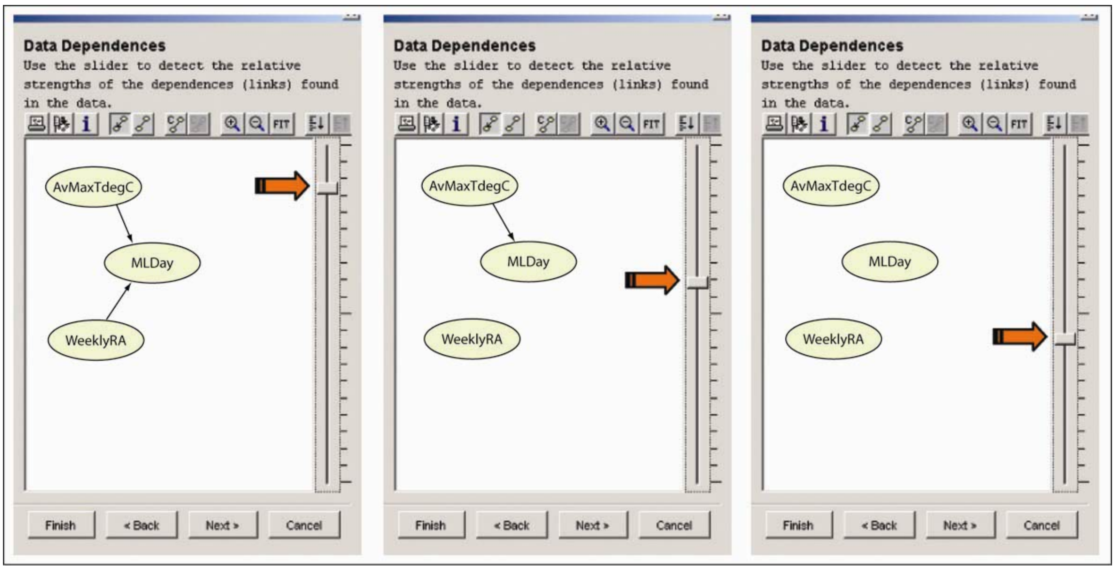


Figure A2.7 Moving the slider on the right downward indicates the relative strength of the links. In this case the link between Temperature and Demand is stronger than between Rainfall and Demand

Step 8: Next, the software provides the opportunity to enter any prior knowledge. In this case we have no idea what the probabilities might be so a ‘1’ is entered for each case to indicate no prior knowledge (Figure A2.8).

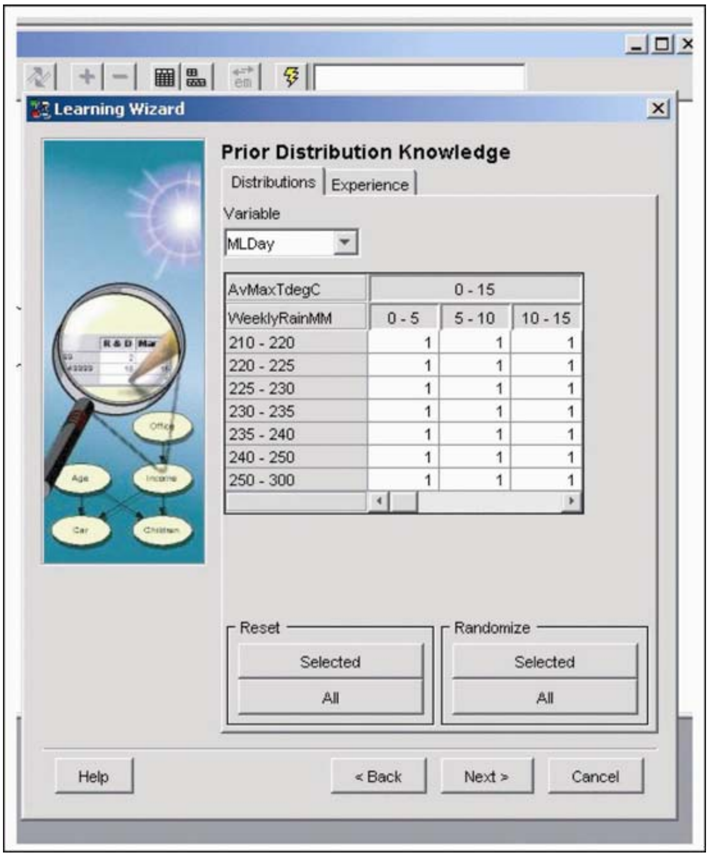


Figure A2.8 CPT for the variable ‘MLDay’ indicating no prior knowledge

Step 9: The final step before calculating the CPT is to specify the number of iterations required for the procedure. Typically a zero is entered here to indicate no upper limit to the number of iterations. The convergence threshold is also required. This must be above zero but is usually taken to be $1x10^{-4}$ (Figure A2.9). For more details of these values and how they are derived the reader is referred to the help menus of the software.

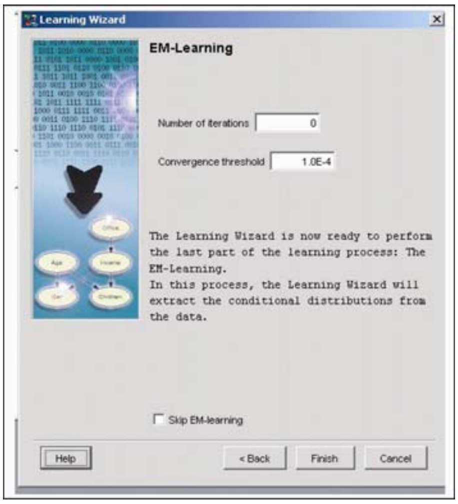
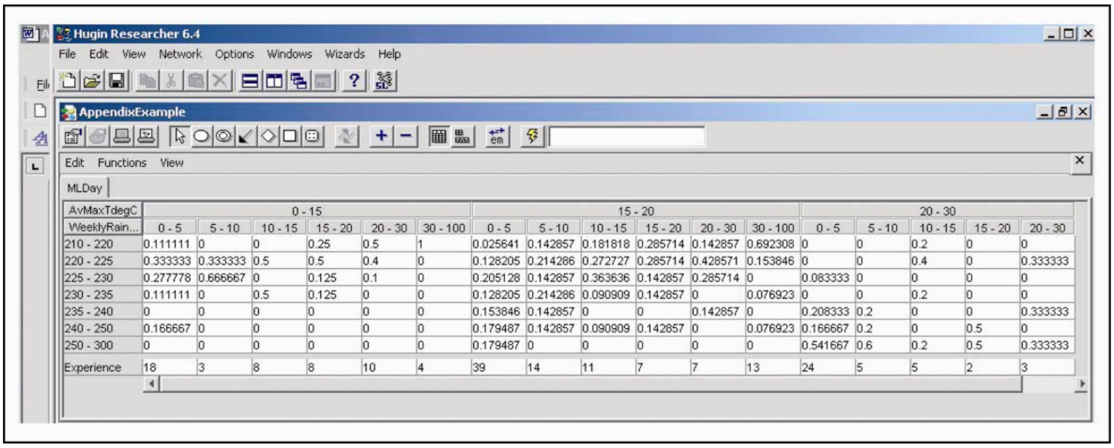


Figure A2.9 Selection of number of iterations and convergence threshold

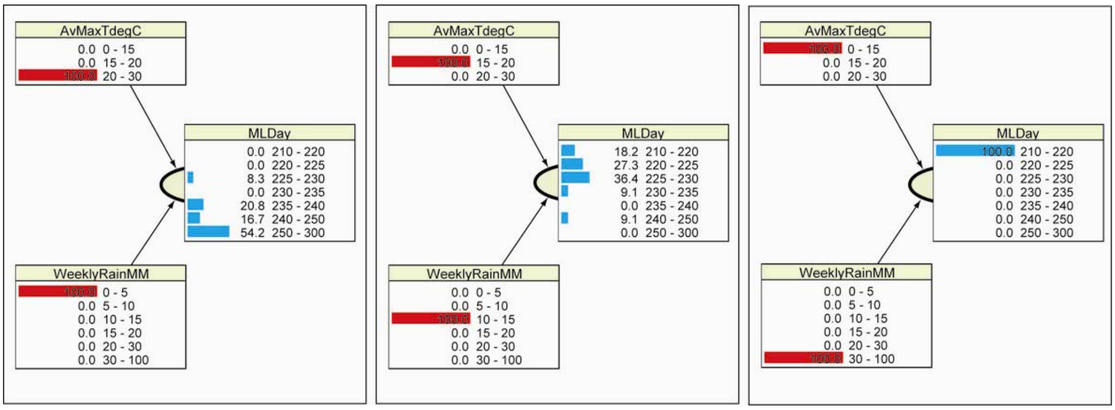
Step 10: The final stage is the generation of the CPTs; the CPT for the variable MLDay is shown in Figure A2.10.



AvMaxTdegC	0 - 15						15 - 20						20 - 30				
WeeklyRain...	0 - 5	5 - 10	10 - 15	15 - 20	20 - 30	30 - 100	0 - 5	5 - 10	10 - 15	15 - 20	20 - 30	30 - 100	0 - 5	5 - 10	10 - 15	15 - 20	20 - 30
210 - 220	0.111111	0	0	0.25	0.5	1	0.025641	0.142857	0.181818	0.285714	0.142857	0.692308	0	0	0.2	0	0
220 - 225	0.333333	0.333333	0.5	0.5	0.4	0	0.128205	0.214286	0.272727	0.285714	0.428571	0.153846	0	0	0.4	0	0.333333
225 - 230	0.277778	0.666667	0	0.125	0.1	0	0.205128	0.142857	0.363636	0.142857	0.285714	0	0.083333	0	0	0	0
230 - 235	0.111111	0	0.5	0.125	0	0	0.128205	0.214286	0.090909	0.142857	0	0.076923	0	0	0.2	0	0
235 - 240	0	0	0	0	0	0	0.153846	0.142857	0	0	0.142857	0	0.208333	0.2	0	0	0.333333
240 - 250	0.166667	0	0	0	0	0	0.179487	0.142857	0.090909	0.142857	0	0.076923	0.166667	0.2	0	0.5	0
250 - 300	0	0	0	0	0	0	0.179487	0	0	0	0	0	0.541667	0.6	0.2	0.5	0.333333
Experience	18	3	8	8	10	4	39	14	11	7	7	13	24	5	5	2	3

Figure A2.10 The final CPT for the MLDay variable. Note the number of 'experiences', or the number of values used to calculate each combination of cases, is displayed in the bottom line. The more experiences, the more reliable will be the probabilities, and the more confidence you can have in the result.

Finally, Figure A2.11 shows the results of the exercise. These examples show that in case (a) with average maximum weekly temperatures between 20-30 degrees and zero rainfall, there is a 54% chance that consumption will be between 250-300 ML day^{-1}. At the other extreme, case (c) indicates that consumption is almost certain to be reduced to between 210-220 ML day^{-1} when conditions are wet and cold. In case (b) with moderate temperatures between 15-20 degrees and a weekly rainfall of 10-15 mm, consumption is most likely to be between 225-230 ML day^{-1}.



(a) Warm and dry (b) Intermediate (c) Cold and wet

Figure A2.11 Some examples of the output using an automatically generated CPT

Appendix III

Completion of Conditional Probability Tables using Stakeholder or Expert opinion

In cases where information is scarce, or when it is difficult to quantify a process, it may be necessary to complete CPTs using input obtained from sources such as stakeholders or experts. There are two types of situation where stakeholder or expert opinion may be sought:

(1) Where local knowledge can substitute for a lack of measurements. This might include historical information about the hydrology, population, incomes etc., which is not recorded, but can be used to help build a CPT. Local residents or organizations may be a source of such information. Stakeholder opinion can also be used to help when trying to quantify social issues, such as the value of an amenity, or acceptability of price changes etc.

(2) Where no data exists, but where expert knowledge of academics, professionals, or others, may be able to provide informed estimates based on theoretical calculations or informed judgement. A good example is provided by the Danish compensation network for which the impacts of different actions on farm economics were quantified using the opinion of a recognized expert in this field.

Both of these information types are subjective and as such open to criticism, but in the absence of data their use is valid, provided the accompanying limitations and uncertainties are acknowledged. A good description of how to complete CPTs using this type of information is given in Cain (2001) .

Cain distinguishes between two situations. The first is where the CPT to be completed is small (less than 10 combinations). In this case the probabilities can be entered directly into the table for all combinations. The second, perhaps more common situation, is where the CPT is larger as a result of either more parents or more states for each of the variables. Remember, a child node with 3 states, having 2 parents each with 3 states, will already have a CPT of 27 combinations. To fill this in directly requires the stakeholder(s) to provide 27 probability estimates, something that is asking a lot of a non-expert; even an expert might find the task a challenge. In this case Cain suggests using an EPT (Elicited Probability Table). In an EPT the probabilities for only a sub-set of the combinations are entered, the rest are calculated using interpolation factors. The procedures are best described using a few simple examples.

Example 1: Direct completion of a small CPT

For this we can use a simplified version of the example in Appendix II. Suppose we want to know how much water is used domestically under different weather conditions, but there is no data. One solution might be to consult an experienced water engineer to provide direct input into our CPT; because we wish to fill in all the combinations, the CPT should be kept as small as possible. Suppose there are 3 variables ‘Temperature’, ‘Rainfall’ and ‘Water Consumption’, with each variable being restricted to two states to minimize the number of combinations. For simplicity the states are described in qualitative terms. The CPT will look like Table A3.1. Because the table is small our engineer will have no problem directly entering the probabilities for each case, based on his experience. Incidentally, it would of course be possible to place threshold values on all the states, so that our definition of ‘Hot’ for instance might be an average temperature above 15°C, and ‘High’ consumption could be above 350 ML day^{-1}.

		‘Water Consumption’	
‘Temperature’	**‘Rainfall’**	*High*	*Low*
Hot	*Dry*	1	0
Hot	*Wet*	0.7	0.3
Cold	*Dry*	0.4	0.6
Cold	*Wet*	0	1

Table A3.1 Example of direct input into a CPT. In this case the stakeholder or expert enters each of the probabilities into the table directly, based on their knowledge and/or experience

Example 2: Completion of an EPT (Elicited Probability Table)

It is not always possible to construct simple CPTs; sometimes they will inevitably be large and complex. In these cases, where entry needs to be manual, direct completion is not practical; instead it is possible to complete a small subset of all combinations and to complete the rest using interpolation factors. Cain terms such tables Elicited Probability Tables (EPTs). As an example let us take the network shown in Figure A3.1.

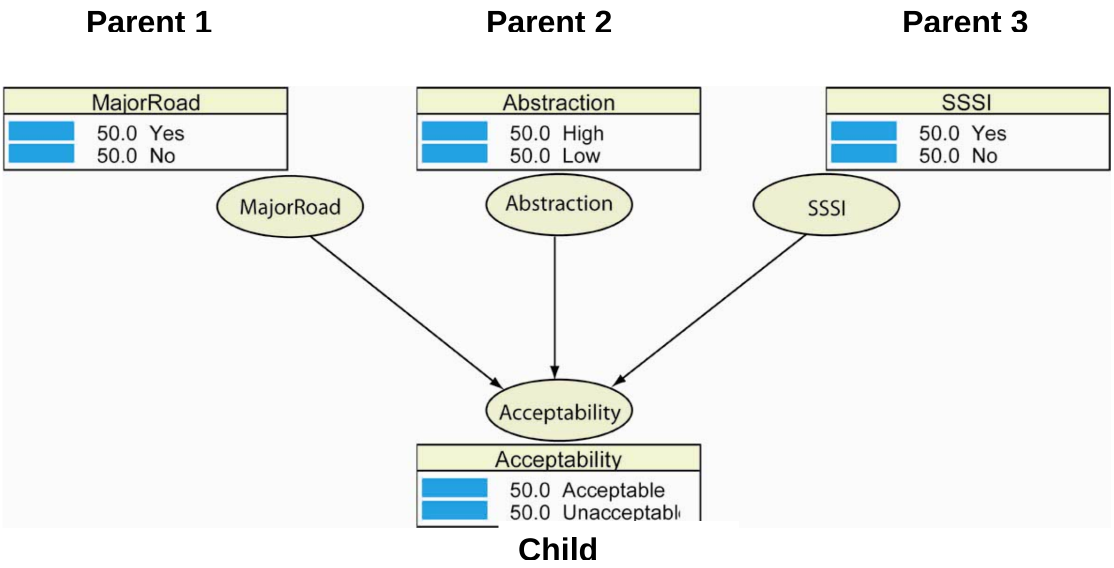


Figure A3.1 Network to illustrate EPT completion

This network investigates the extent to which development of an amenity area (perhaps an area of natural beauty) is acceptable to the stakeholders. In this example the impact of three possible actions are considered; the construction of a major road, abstraction from a river, and the declaration of the site as a SSSI (site of special scientific interest). All the variables are given two states, and it is assumed that the three parents do not affect the degree of change that the others have on the child (i.e. they are NOT modifying parents: see Step 3.3 in the guidelines).

The child has two states, one of which (Acceptable) is more desirable than the other; this can be termed the “success” state. Each parent is also taken to have two discrete states, one of which is more likely to give rise to the success state of the child node. This can be called the “positive” state. For example, for Parent 1 (Road

Construction) the state ‘No’ is positive from the point of view of improving the site as a public amenity, whereas ‘Yes’ would be negative, since this might be considered to detract from the area.

When the design of the network is complete, and we are ready to complete the CPTs, a table should be drawn up as shown in Table A3.2. These represent questions that can be put to the stakeholders.

	Parent 1	**Parent 2**	**Parent 3**	**Child State: score out of 10**
Question 1:	Positive state of P1 *(No)*	Positive state of P2 *(Low)*	Positive state of P3 *(Yes)*	Elicited state & score out of ten
Question 2:	Negative state of P1 *(Yes)*	Negative state of P2 *(High)*	Negative state of P3 *(No)*	Elicited state & score out of ten
Question 3:	Negative state of P1 *(Yes)*	Positive state of P2 *(Low)*	Positive state of P3 *(Yes)*	Elicited state & score out of ten
Question 4:	Positive state of P1 *(No)*	Negative state of P2 *(High)*	Positive state of P3 *(Yes)*	Elicited state & score out of ten
Question 5:	Positive state of P1 *(No)*	Positive state of P2 *(Low)*	Negative state of P3 *(No)*	Elicited state & score out of ten

Table A3.2: Elicited probability table (EPT) for network in Figure A3.1

The questions represent only a sub-set of the 16 possible combinations that exist in the CPT (i.e. 2 x 2 x 2 x 2 combinations). They have been chosen to cover the extreme range of probabilities that might be encountered in the CPT, and be the easiest for the stakeholder(s) to answer.

The EPT is formally structured, as follows. The first line (question 1) is such that the parents are all in their positive states. This question asks the stakeholder what state they consider the child would be in, if all the parent nodes were in their positive state. In the second line (question 2) the parents are all in their negative states. For all the other lines, each parent in turn is “switched” from its positive state to its negative state. This is done one parent at a time so that, after the first two, each line only ever has one state that is negative.

Before beginning the interview, you should run through, with the stakeholder, the sort of questions you are going to ask. Explain that you will ask the questions in sequences with each sequence being linked to the same child node. For the first question, ask the stakeholder to imagine that a major road is not being constructed in the area, (positive P1), that there will be no abstractions from the river (positive P2), and that the region will be protected by being declared a SSSI (Positive P3). Ask the stakeholder to give a score out of 10 for the acceptability of this set of actions, with 10/10 indicating total acceptance. The likelihood is that the response to this will be a score of 10/10 for the success state (Acceptable) of the child. The response to the next question (all negative states for the parents) is likely to elicit a 0/10 score. Succeeding questions will provide intermediate scores depending on the importance the stakeholder attaches to each parent.

These questions can be posed to a single representative stakeholder, or preferably to a large number. The responses from each can then be averaged to give the final table.

Let us assume that the stakeholders provide the responses shown in Table A3.3; this leaves us to interpolate the scores for the remaining 11 combinations.

	Parent 1	Parent 2	Parent 3	Child State: score out of 10 that child is in positive state
Question 1:	Positive state of P1 *(No)*	Positive state of P2 *(Low)*	Positive state of P3 *(Yes)*	10
Question 2:	Negative state of P1 *(Yes)*	Negative state of P2 *(High)*	Negative state of P3 *(No)*	0
Question 3:	Negative state of P1 *(Yes)*	Positive state of P2 *(Low)*	Positive state of P3 *(Yes)*	6
Question 4:	Positive state of P1 *(No)*	Negative state of P2 *(High)*	Positive state of P3 *(Yes)*	7
Question 5:	Positive state of P1 *(No)*	Positive state of P2 *(Low)*	Negative state of P3 *(No)*	9

Table A3.3: Stakeholder responses to the EPT

Table A3.4 shows all the combinations of the three parents, including those for which no values have been provided by the stakeholders (i.e. Combination states 4, 6 and 7). These can be calculated using interpolation factors.

State combination	Parent 1	Parent 2	Parent 3	Child State: score out of 10 that child is in positive state
1	Positive *(No)*	Positive *(Low)*	Positive *(Yes)*	10
2	Positive *(No)*	Positive *(Low)*	Negative *(No)*	9
3	Positive *(No)*	Negative *(High)*	Positive *(Yes)*	7
4	**Positive *(No)***	**Negative *(High)***	**Negative *(No)***	**?**
5	Negative *(Yes)*	Positive *(Low)*	Positive *(Yes)*	6
6	**Negative *(Yes)***	**Positive *(Low)***	**Negative *(No)***	**?**
7	**Negative *(Yes)***	**Negative *(High)***	**Positive *(Yes)***	**?**
8	Negative *(Yes)*	Negative *(High)*	Negative *(No)*	0

Table A3.4 The complete CPT showing missing values

Interpolation factors are obtained in the following way:

Interpolation factors are obtained for each 'switch' in the state of a parent from positive to negative. They are calculated in relation to the difference between the highest probability (all parents in the positive state), and the lowest (all parents in the negative state). Using the combination numbers in Table A3.4, this can be expressed as P_1-P_8. When one of the parents is switched from a positive to a negative state, the probability of the child being in the success state is reduced. The interpolation factor simply quantifies this reduction, for each parent, as a proportion of P_1-P_8. Thus:

Interpolation Factor (IF) for Parents

IF3 (Parent 3) = $(P_2-P_8) / (P_1-P_8)$ = (9-0)/(10-0) = 0.9
IF2 (Parent 2) = $(P_3-P_8) / (P_1-P_8)$ = (7-0)/(10-0) = 0.7
IF1 (Parent 1) = $(P_5-P_8) / (P_1-P_8)$ = (6-0)/(10-0) = 0.6

Interpolation factors calculate the way the probability of the state of a child changes when a parent switches from a positive to negative state. Thus, to calculate the child state in Table A3.4 for combination state 4, which has not been given by the stakeholders, all we need to do is to multiply state combination 3, by the interpolation factor associated with parent 3 (i.e. IF3). The only difference between state 3 and state 4 is that parent 3 has switched from positive to negative.

Thus for the three unknown probabilities:

$P_4 = [(P_3-P_8) \times IF_3] + P_8 = [(0.7 - 0) \times 0.9] + 0 = 0.63$
$P_6 = [(P_5-P_8) \times IF_3] + P_8 = [(0.6 - 0) \times 0.9] + 0 = 0.54$
$P_7 = [(P_5-P_8) \times IF_2] + P_8 = [(0.6 - 0) \times 0.7] + 0 = 0.42$

When these interpolation factors are applied, the resulting CPT is shown in Table A3.5

State combination	**Parent 1**	**Parent 2**	**Parent 3**	**Child State: score out of 10 that child is in positive state (Acceptable)**
1	Positive *(No)*	Positive *(Low)*	Positive *(Yes)*	10
2	Positive *(No)*	Positive *(Low)*	Negative *(No)*	9
3	Positive *(No)*	Negative *(High)*	Positive *(Yes)*	7
4	**Positive *(No)***	**Negative *(High)***	**Negative *(No)***	**6.3**
5	Negative *(Yes)*	Positive *(Low)*	Positive *(Yes)*	6
6	**Negative *(Yes)***	**Positive *(Low)***	**Negative *(No)***	**5.4**
7	**Negative *(Yes)***	**Negative *(High)***	**Positive *(Yes)***	**4.2**
8	Negative *(Yes)*	Negative *(High)*	Negative *(No)*	0

Table A3.5 The CPT with the calculated probabilities for the child being in an Acceptable state shown in red. The probabilities are given as a score out of 10.

In Table A3.5 the figures in the right hand column show the probability that the child variable is the positive state (e.g. for state combination 6 the probability of the combination being acceptable to stakeholders is 54%). Table A3.5 only shows the 8 probabilities that indicate the child is in the positive (Acceptable) state. The remaining 8 probabilities, that the child is in the negative (Unacceptable) state, is obtained by simply subtracting the positive values from 10. The final appearance of the CPT in the network, with all 16 combinations completed, is shown in Table A3.6.

Acceptance	**Road (1)**	Yes				No			
	River (2)	High		Low		High		Low	
	SSSI (3)	Yes	No	Yes	No	Yes	No	Yes	No
Yes		**6**	**0**	**4.2**	**5.4**	**7**	**6.3**	**10**	**9**
No		**4**	**10**	**5.8**	**4.6**	**3**	**3.7**	**0**	**1**

Table A3.6 Completed CPT for the variable 'Acceptability'; note all columns add up to 10

In his appendices Cain goes on to describe how to construct EPTs for a range of different circumstances including:

1. When one or more of the parents is a modifying variable
2. When one or more parent is a continuous variable
3. When the child has 3 or more states

Readers are referred to these appendices in Cain's report, which is included on the CD attached to these guidelines.

To facilitate the completion of EPT tables Cain also developed a calculator, which he has kindly made freely available, and which we have also included on the attached CD.

References

Cain, J., 2001. Planning improvements in natural resources management. Centre for Ecology and Hydrology, Wallingford, UK, 124pp pp.