International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

A Belief Network Approach to Optimization and Parameter Estimation in Resource and Environmental Management Models

Varis, O.

IIASA Working Paper

WP-95-011

February 1995

Working Paper

A Belief Network Approach to Optimization and Parameter Estimation in Resource and Environmental Management Models

Olli Varis

WP-95-11
February 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

A Belief Network Approach to Optimization and Parameter Estimation in Resource and Environmental Management Models

Olli Varis

WP-95-11
February 1995

IIASA International Institute for Applied Systems Analysis ◻ A-2361 Laxenburg ◻ Austria
Telephone: +43 2236 807 ◻ Fax: +43 2236 71313 ◻ E-Mail: info@iiasa.ac.at

Preface

Understanding and accounting for uncertainty is extremely important in environmental management. Unlike electrical or mechanical engineering where most outcomes can be computed, natural systems often react to human interference in unexpected and unpredictable ways. Data uncertainty is in many cases only part of the reason, the other being lack of full understanding and knowledge about the processes generating response in natural systems such as rivers and lakes. Therefore mathematically handling uncertainty was, and still is, a major challenge in water quality management.

Several mathematical theories have been developed and used in uncertainty calculus. Some are based on probability concepts, others use fuzzy set theory, still others use axiomatization of upper and lower probabilities to extend Bayesian inference rules (Dempster-Schaffer theory of evidence). It is not yet clear which methodology is superior and what are the merits and shortcomings of each of them from the practical point of view because real-life applications and case studies are still lacking. The present work introduces an attempt at yet another axiomatization for handling uncertainty, based on the author's extension to existing theory of Bayesian belief networks. The theoretical presentation is followed by an illustration in the application to a river water quality management problem.

The methodology proposed by the author is quite new and so far received little testing in either applied or theoretical studies. It is rather an invitation for both experts in mathematical theory of decision making and for water quality specialists to discuss and evaluate the practical feasibility and theoretical grounds of this technique and to help find its proper place among well-known and emerging instruments of analysis in decision making field.

Abstract

This study presents an approach to use Bayesian belief networks in various optimization tasks in resource and environmental management. A belief network is constructed to work parallel to a deterministic model, and it is used to update conditional probabilities associated with different components of the model. The propagation of probabilistic information occurs in two directions in the network. The divergence between prior and posterior probability distributions at model components can be used as indication on inconsistency between model structure, parameter values, and other information used. An iteration scheme was developed to force prior and posterior distributions to become equal. This removes inconsistencies between different sources of information. The scheme can be used in different optimization tasks including parameter estimation and optimization between various management alternatives. Also multiobjective optimization is possible. The approach is illustrated with two numerical examples and with a hypothetical example on cost-effective management of river water quality.

Table of Contents

1.	**Introduction**		**1**
2.	**The Uncertainty Balance Approach**		**4**
	2.1	*Structure, targets, and uncertainty balance*	*4*
	2.2	*Non-informative network implies full balance*	*4*
	2.3	*New information induces a need to re-establish the balance using control variables and network links*	*5*
	2.4	*Iteration for balance*	*6*
3.	**Computational Solution**		**7**
	3.1	*Propagation in outcome layer (deterministic equations)*	*7*
	3.2	*Information from outcome layer to probabilistic layer*	*8*
	3.3	*Propagation in probabilistic layer*	*10*
	3.4	*Information from probabilistic layer to outcome layer*	*13*
4.	**An Example on Two-Directional Propagation**		**14**
5.	**An Example on Parameter Estimation by Uncertainty Balance**		**18**
	5.1	*The model without external information*	*18*
	5.2	*One observation is included*	*19*
	5.3	*More than one observation*	*19*
	5.4	*Parameter estimation*	*20*
	5.5	*Simplifying the configuration of the probabilistic layer*	*21*
6.	**An Example on River Water Quality Management**		**23**
	6.1	*The management problem and the watershed*	*23*
	6.2	*A probabilistic river model*	*23*
	6.3	*Model vs. data: illustration of the approach*	*25*
	6.4	*River water quality management scenarios*	*30*
7.	**Discussion and Conclusions**		**35**
Acknowledgments			**36**
References			**37**
Appendix 1			
Appendix 2			
Appendix 3			

A Belief Network Approach to Optimization and Parameter Estimation in Resource and Environmental Management Models

OLLI VARIS
Helsinki University of Technology, Laboratory of Hydrology and Water Resources Management, L-Building, FIN-02150 Espoo, Finland
ovaris@leka.hut.fi

1. Introduction

Uncertainty is evidently among the most discussed topics in environmental and resource management. Interest in probabilistic assessment, risk analysis, and related techniques has grown rapidly in the recent years (WCED 1987, De Jongh 1988, USEPA 1989, ADB 1990). Probabilistic and risk analysis types of approach are increasingly accepted in practical assessment work by international organizations and by national authorities in several countries. Modern decision theory, together with various recently developed computational techniques for processing uncertain information, provide a wide base for novel, potential approaches to applications in the field. At present, these opportunities are far from being properly known and fully utilized.

The concept of uncertainty has several facets in this context. From a decision-theoretic view, uncertainty can be grouped in three clusters (cf., Howard 1968, Varis et al. 1994):

- Acquisition, presentation, and propagation of information available.
- Preferences and objectives of a given problem.
- Structural issues.

Pearl (1988) divides the computational techniques used into two groups: logic-based approaches (monotonous logic and its applications; for example, rule-based systems, etc.) and probabilistic approaches (Bayes, Dempster-Shafer, fuzzy set theory, etc.). In this study, the focus is on Bayesian calculus due to its many favorable properties (Howson & Urbach 1991).

Within environmental and resource management, the applications of Bayesian analysis have been largely dominated by classical Bayesian inference, i.e., parameter estimation, in which the Bayesian analysis is restricted to the parameter space. In decision theory, the idea of considering the entire model as a construct subject to uncertainty and subjectivity stem from the game theory of the 1930s and '40s (Shafer 1990). Games evolved into sequential games against uncontrolled 'nature', and abstractions such as decision trees were developed. Bayesian decision theory – in contrast to Bayesian inference – gained increasing notice and emphasis (Wald 1950). These theoretical concepts were developed into more applicable ones until the late 1960s (Howard 1968, North 1968, Raiffa 1968).

Further development has been linked with advances in related computational mathematics (Shafer & Pearl 1990). Artificial intelligence has had a rapidly growing impact within the last ten years. A set of probabilistic, Bayesian-type approaches applicable or potentially applicable to decision analysis under high uncertainty has emerged (Horwitz et al. 1988, Pearl 1988, Shafer 1990, Szolovits & Pauker 1993). Characteristic of these techniques — known as belief networks, causal networks, Bayesian nets, qualitative Markov networks, or constraint networks — is the principle of networking nodes representing conditional, locally updated probabilities. The local-updating principle allows construction of large and densely coupled networks without excessive growth in computation. Furthermore, networks can easily be constructed to operate interactively and on-line. In recent years, they have spread quickly to many application areas, including fault diagnosis, reliability theory, medicine, and pattern recognition.

According to Bobrow (1993), a particularly successful networking technique has been the belief network approach by Pearl (1986a, 1988), which was also used in the present study. A review by Szolovits & Pauker (1993) stated that '*... Pearl's formulation has had a revolutionary impact on much of A[rtificial] I[ntelligence]*'. As is usual in such techniques, the entire model – *the hypothesis space* – is subjected to Bayesian analysis, not only the parameter space (cf. Gordon & Shortliffe 1985, Pearl 1986b, Shenoy & Shafer 1986). In contrast to classical probability theory, different sets of outcomes are allowed for related nodes, yielding an evident violation of the Kolmogorov axiomatization of the Bayes formula, yet Pearl (1986a) strongly argues against this very axiomatization: '*It is not hard to see that this textbook view of probability theory presents a rather distorted view of human reasoning and misses its most interesting aspects.*' Many decision analytic approaches have also been in line with these ideas (see, for instance, Shachter 1986, 1988).

A recent study by Varis (1994) examined the belief network methodology of Pearl (1986a, 1988), and offered suggestions for making the approach more suitable for decision analysis in resource management and environmental studies. The suggested methodology has borrowed ideas particularly from Bayesian decision analysis and from some common practices within the field. The most essential suggestion was that nodes can be linked in two layers: (1) the probabilities can be propagated using belief-function calculus; and (2) the outcome values of the nodes can be linked using deterministic equations (algebraic or logical). This implies that the network is understood as an approximate, numerical approach to updating the uncertainty in different parts of the model, making probabilistic simulations (such as Monte Carlo analysis) useless. This updating works instantly and does not require off-line simulation runs; in other words, uncertainty is not considered as an externality, introduced off-line to certain model components such as parameters, as in Monte Carlo analysis, but rather it is included as an intrinsic property of the structure of the model.

As far as the handling of uncertainty is considered, all three categories of uncertainty listed above (i.e., propagation and presentation, objectives and preferences, and structure) are supported. Uncertain information is propagated using discrete belief-function calculus. A special, very valuable property of the approach is the ability to submit information in two directions, i.e., from variable A to variable B, and vice versa. Our information on A can be used to update our information on B, and vice versa. As far as the presentation and analysis of uncertainties in objectives and preferences are concerned, the discrete probabilistic domain allows the use of many of the basic concepts of expected utility theory, including risk-attitude analysis and value-of-information analysis. Structural uncertainty is handled in the two-layered network in the following manner: above, the variables can be linked by deterministic equations, and below by a network of conditional probabilities. This structure allows a degree of belief to be assigned to a deterministic dependency between variables. The approach was also shown to facilitate a number of possibilities to use model elements from different modeling traditions.

The goal of this study was to formulate and test the use of this approach in optimization and parameter estimation. The basic concept — stemming from the idea of using a belief network 'below' a deterministic model in to handle uncertainties in different parts of the model — is to look at inconsistencies between the *model* outcomes and *external information* such as management targets (cost levels, environmental indices, etc.) or observations to which the model should be fitted. Inconsistencies are shown by diverging prior and posterior probabilities in certain parts of the model. An iteration scheme was developed to adjust the model using two types of variables to attain consistency between various information sources. These two variable types are control variables (such as parameters or wastewater treatment levels) and network link properties.

The approach is illustrated by two simple numerical examples, and with a more extensive example of a water quality management model of a river basin. The hypothetical river example is based on recent river water quality studies for Central and Eastern European countries (e.g., Somlyódy et al. 1994). The problem is to find cost-effective wastewater treatment solutions on a river basin scale.

2 The Uncertainty Balance Approach

Assume that we must solve a complex control and/or diagnosis problem with high uncertainties. The available information comes from diverse sources and is contradictory. We need a balanced view of the problem based on all the information sources.

2.1 Structure, targets, and uncertainty balance

Assume that we have or construct *a model* to describe the crucial elements of the problem. We want to use the structure provided by the model as a basis for our reasoning. In addition, we have *information that is external to the model* (knowledge, experience, data, goals, etc.). All the information is uncertain. We want to put our diverse, uncertain, and contradictory information into an analytic framework in which a reasonable compromise and balance between different pieces of information can be found.

In the approach presented in this study, the model is deterministic, relatively simple management oriented tool. There are observations, target levels, and/or some other pieces of external information that can be used together with the model. Technically, the approach divides the model into two layers in communication with one another. The deterministic equations of the model constitute *the outcome layer*. The term 'outcome' is used because from that layer, one can get numerical values for the model variables (e.g., oxygen levels in a river). The other, *probabilistic layer*, consists of a network of approximate, conditional distributions for different outcome values of the model. The network is based on belief network methodology (Pearl 1988, Varis 1994).

The terms 'outcome' and 'associated probability' can be somewhat confusing. The following simple example should clarify these terms. A baby is to be born. There is prior evidence of an equal probability that the baby will be either a boy or a girl. Formally:

$$P(boy) = 0.5$$
$$P(girl) = 0.5$$

Now, *boy* and *girl* are outcomes, and 0.5 and 0.5 are their probabilities, respectively. These probabilities are often called evidence, since they typically are based on experience, data, logical reasoning, or other information giving evidence of reasoning and which is external to the model used.

2.2 Non-informative network implies full balance

Now we return to the two-layered model framework. The major issue of this study is how to use belief networks to assist in parameter estimation and, more generally, when optimizing other control variables to fulfill the targets defined. The key proposition is that the prior and posterior probability distributions of the target variables (observations, management targets, constraints, etc.) should become equal. This implies that the joint distributions of the external information (prior) should be equal to the modeled distributions of these variables (posterior) and assures that the prior information is properly utilized in the analysis. The above is done by finding optimal values for the control variables (including parameters) by iteration.

The belief network is constructed so that if there is no target information diverging from the model prediction, then all the discrete probability distributions in the network are uniform distributions. In Bayesian terminology, a non-informative prior probability distribution of a quantity is defined as a distribution in which there is an equal probability of occurrence for each possible outcome of the variable. Clearly, this is a rather contradictory concept because a uniform distribution is as informative as any other probability distribution. In the approach proposed, a model with no external information defines a probabilistic layer in which all probability distributions are uniform. These uniform distributions now, indeed, define an exact probability distribution — namely the distribution of the corresponding variable in the outcome layer — and the above-mentioned problem is avoided.

A discrete approximation of a distribution in the outcome layer is made, and a probability distribution with a number of outcomes with equal probabilities of occurrence is obtained. In the present study, three outcomes are used, but without the loss of generality the number need not be three. The concept of three-point approximation of continuous variables is well-studied and widely applied (Keefer and Bodily 1983, Miller and Rice 1983), and the benefits of using a higher number of outcomes — giving slightly more accuracy to computations — should be compared with the increased computational effort needed and the accuracy of the external information available. Evidently, the biggest problems due to inaccurate approximations are the tails of the distributions. The more extreme events are under consideration; the more important is the role of these inaccuracies.

In summation, if the probabilistic layer consists of uniform distributions only, this tells us that no information is available other than that provided by the model, or, if there is external information, it is in full agreement with the model.

2.3 New information induces a need to re-establish the balance using control variables and network links

Include now a piece of external, probabilistic information (target: observation, management goal, etc.) in the analysis. Its probability distribution is approximated with a discrete distribution in which the outcome values are the same as in the corresponding model component, but in general the probability values become different. As an example, assume that an experiment at a hospital has shown a probability of 0.8 that the baby will be a boy. Now, the outcomes have remained the same (*boy, girl*), but their probabilities have changed.

The probabilistic layer is used to propagate this new information throughout the model. Evidently, all distributions deviating from the uniform distribution indicate that the model and the external information do not match completely. A controversy exists and it needs to be analyzed, and a proper balance should be found.

This can be done by looking more carefully at certain parts of the model, these parts are the variables with which the model can be controlled, i.e., the decision/control variables. They can be, for instance, model parameters used to fit the model to data (match targets), or wastewater treatment plants along a river which can be upgraded to various purification levels to control (improve) water quality in the river (again a target). In the latter, another set of targets may be due to the costs involved, and a balance fulfilling these targets should be found.

According to the proposition made in the previous section, the balance can be found by forcing the distributions calculated by the probabilistic layer (posterior distributions) to be uniform. This implies that the joint distributions of the external information are equal to that of the modeled information. This can be achieved by changing the outcome distributions of the control variables under consideration, until this goal is attained. The form of the posterior distribution gives a clear indication of how these distributions can be found. Another set of components that can be controlled to achieve the balance are the parameters describing how strongly two variables in the model are interlinked. If, for instance, a link strength corresponding to a deterministic model equation = 1, then we assume that this equation is 100% adequate in describing the phenomenon it should describe. If the link strength = 0, then we assume that the equation tells nothing on the phenomenon. Moreover, these link strengths clearly influence the model uncertainty calculated at the outcome layer. The lower the link strength is, the further the error bounds are from the expected behavior of the system. The reason for this is that link strengths enable us to take into account the structural uncertainty of the model.

2.4 Iteration for balance

We want to achieve a situation in which the joint probability distributions of all probabilistic, external information propagated into control variables equal the prior distributions of these variables. The search is done by iteration, and the probabilistic layer is used as a numerical solution to that problem. Figure 1 shows the outline of the uncertainty balance iteration. An intrinsic component of the analysis is the analyst her/himself, because much of the benefit of such analyses in non-trivial problems comes from the learning from and interaction with the information available. Therefore, the approach has been designed to be as interactive as possible and to be operated on-line.

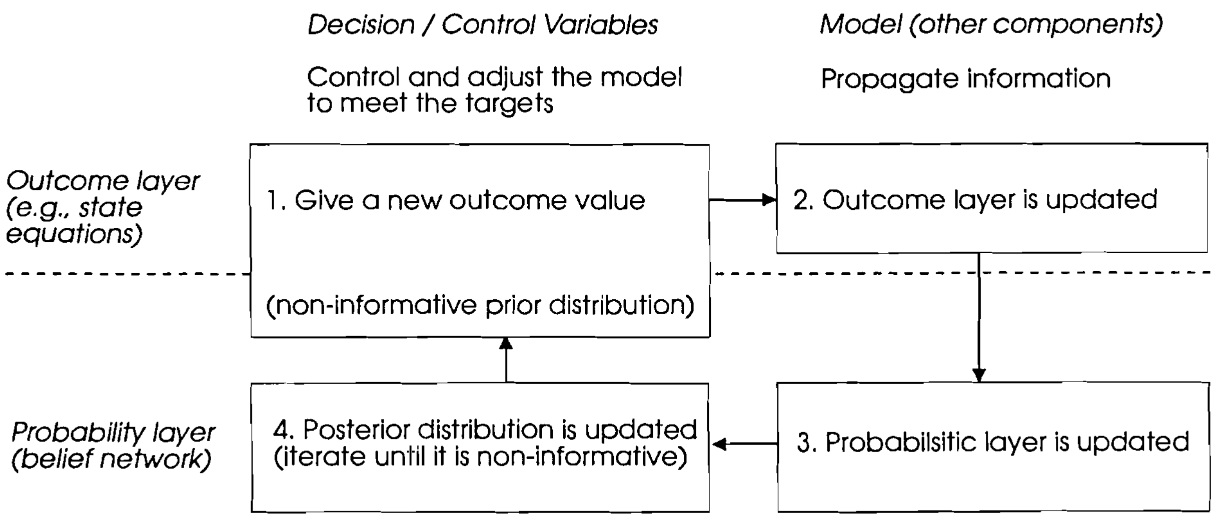


Figure 1. Outline of the uncertainty balance iteration.

3. Computational Solution

The procedure outlined above is described in detail below as four subsequent steps. A list of repeatedly used mathematical symbols is given in Appendix 1.

3.1 Propagation in outcome layer (deterministic equations)

If the state equations are nonlinear, as is very often the case in practice, the analytical propagation of uncertainty is usually too laborious. There is a myriad of approximate approaches to propagation of uncertainty in deterministic equations (see, for instance, Korn and Korn 1968, Morgan and Henrion 1990). One of the most widely used approaches is the Taylor series expansion. The more accuracy required for the approximation, the more terms can be included in the analysis. We consider here the first-order approximation, which, in many cases (such as those shown in the examples later) is sufficiently accurate. For equations expressing the deviations in output y from its nominal value, caused by deviations of $x_1,\ldots x_n$ from their nominal values, the first-order approximation for the variance of y is

$$\mathrm{var}[y] \approx \sum_{i=1}^{n} \mathrm{var}[x_i] \left[\frac{\partial y}{\partial x_i} \right]^2_{X^0} \qquad (1)$$

There are two specific cases in which rather practical equations for expected value and uncertainty of y can be derived: the weighted sums of components and products of powers of the components. In the case of weighted sums

$$y = \sum_{i=1}^{n} a_i x_i \qquad (2)$$

the mean and the variance can be obtained by

$$E[y] = \sum_{i=1}^{n} a_i E[x_i] \qquad (3)$$

$$\mathrm{var}[y] = \sum_{i=1}^{n} a_i^2 \,\mathrm{var}[x_i] + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j \,\mathrm{cov}[x_i, x_j] \qquad (4)$$

Accordingly, for product and power equations

$$y = \prod_{i=1}^{n} x_i^{a_i} \qquad (5)$$

the mean and the variance are

$$E[y] \approx \prod_{i=1}^{n} E[x_i]^{a_i} \qquad (6)$$

$$\mathrm{var}[y] \approx \sum_{i=1}^{n} \mathrm{var}[x_i]\left(\frac{a_i}{x_i^0} y^0\right)^2 \qquad (7)$$

The variance equation can be processed in a more convenient form by using the coefficient of variation

$$\mathrm{cv}(z)^2 \equiv \frac{\mathrm{var}[z]}{E^2[z]}$$

hence,

$$\mathrm{cv}^2[y] \approx \sum_{i=1}^{n} a_i^2 \mathrm{cv}^2[x_i] \qquad (8)$$

Often, the state equations of models, especially when they are analytical solutions of differential equations, contain an exponential function, whose uncertainty can be propagated in the following approximate manner. The uncertainty of the exponent can be obtained by Equation (5). Denote this augmented exponential function as

$$x_i = e^{-x_i^*} \qquad (9)$$

Using now the general, first-order approximation equation (7), we obtain

$$\mathrm{var}[x_i] \approx \mathrm{var}[x_i^*] e^{-x_i^{*0}} \qquad (10)$$

Since $e^{-x_i^{*0}}$ is close to 1, but always below it if the sign of the exponent is negative, we are safe in approximating the variance as

$$\mathrm{var}[x_i] \approx \mathrm{var}[x_i^*] \qquad (11)$$

Above, it was assumed that the model is structurally correct. In the present approach this does not need to be the case. As will be shown later, an uncertainty estimate can be given of the model structure, expressed as link strength η, a parameter defining a link matrix in the probabilistic layer. Details are given in step 3.3. The link strength can be augmented to the outcome layer in the following approximate manner:

$$\mathrm{cv}'[y] \approx \frac{\mathrm{cv}[y]}{\sqrt{\eta}} \qquad (12)$$

where cv'[y] is the coefficient of variation of the model prediction when the model structural uncertainty is included. In cv[y] it is excluded.

3.2 Information from outcome layer to probabilistic layer

As was discussed in Section 2.2, the terminology of Bayesian statistics refers to a uniform distribution used as the prior distribution as a non-informative prior. It is well known that this distribution describes an exact probability distribution. The use of non-informative priors has been widely criticized, and ways of avoiding the need to use an exactly defined probability distribution as the prior have been developed. One such is the Dempster-Shafer approach (Caselton & Luo 1992).

In this study, uniform prior probability distributions are used. Yet, they are not defined as non-informative, but instead represent an exact distribution as described below, thus avoiding the problem mentioned above. The model prediction y_i at point i is a normally distributed variable (Figure 2), whose probability distribution is approximated by a discrete one with n equally likely intervals. Hence, in a network with no measured information, all distributions are uniformly distributed. This does not mean that there is no information, as the Bayesian concept of non-informative priors would suggest, but instead, implies, that the net contains no information that would contradict the information propagated by the outcome layer of the model. If any external information (measured information, target level, etc.) differing from the model prediction is included, then non-uniform distributions are indications of it in the net.

A further, practical rationale for using uniform distributions in the probabilistic layer in the sense mentioned above is that a vector product of two discrete uniform distributions is a uniform distribution. This feature is important when propagating information in the probabilistic layer, as will be shown later.

Since the outcome layer uses continuous distributions and the probabilistic layer is in discrete form, discrete approximations of the continuous random variables are needed when taking them as priors to the probabilistic layer. The following approximation is used.

First, define y_1 and y_2 such that

$$\begin{aligned} &P(y \leq y_1) = 1/3, \\ &P(y_1 < y \leq y_2) = 1/3 \\ &P(y > y_2) = 1/3 \end{aligned} \quad (13)$$

These values can be obtained by, e.g., using standard normal deviates (for instance, z-tables from statistics):

$$\begin{aligned} y_1 &= \mu_y - 0.4307\ \sigma_y \\ y_2 &= \mu_y + 0.4307\ \sigma_y \end{aligned} \quad (14)$$

In other words, the model prediction is approximated with a discrete distribution with three equally likely intervals. These values can then be used to find the discrete approximation $\mathbf{e}$ for e^*. This will now be made using the intervals obtained above.

$$\begin{aligned} P(e_1) &= P(e \leq y_1) \\ P(e_2) &= P(y_1 < e \leq y_2) \\ P(e_3) &= P(e > y_2) \end{aligned} \quad (15)$$

These values are used as the evidence vector in the probabilistic layer.

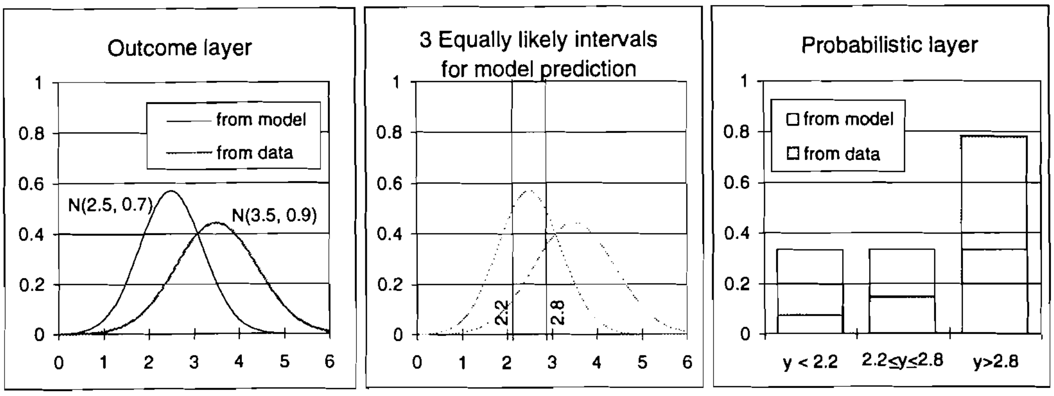


Figure 2. Discrete approximation of an observation.

3.3 Propagation in probabilistic layer (belief network)

In the propagation of information in the probabilistic layer, the belief network approach by Varis (1994) is used. It is based on the approach by Pearl (1988) with a set of modest extensions. Varis (1994) provides some examples and more details of the approach itself and its potential applicability in environmental and resource management modeling. Here, an outline is given.

The probabilistic layer (belief network) consists of nodes connected with links. Those properties of nodes, links, and networks that are relevant to this study are described.

Nodes. Each node i in a network contains

- A vector of possible (discrete) outcomes $\mathbf{y}_i$ that can be defined as inputs, or they may depend on the outcome values of other nodes.
- An evidence vector $\mathbf{e}_i$, with probabilities $e_1,...,e_k$ assigned to k outcomes. In the present study, the number of outcomes is three. The evidence vector transmits external information (data, targets, etc.) to the model.
- A posterior probability distribution $\mathbf{Bel}_i$.

The prior probabilities assigned to the outcomes are updated with information linked from other parts of the network, yielding the posterior probability distribution.

Links. A probabilistic link (uncertainty link) transfers information from one node to another. It is defined as the link matrix $\mathbf{M}_{i|j}$ between two variables i and j, denoting the conditional probability of i given j. In the simplest case of a unidirectional chain, the link matrix equals a Markov chain state transition matrix.

Since the probabilistic layer parallel to the deterministic equations describes their structural uncertainty, the distribution of i should have all the other properties of the distribution of j except variance, which is changed intentionally. These other properties are the moments of the distribution: expected value, skewness, and kurtosis. The variance should be increased correspondingly to the amount of structural uncertainty. It is often practical, for instance in the present study, to give the strength of each link using a single parameter instead of inserting values for each link matrix component separately. The following approach, as shown in Appendix 3, fulfills the moment requirements stated above.

The link strength parameter is denoted as $\eta_{j|i}$, $i \neq j$. $\eta_{j|i} \in [-1, 1]$. A symmetric, k x k link matrix $\mathbf{M}_{j|i}$ is constructed as a function of $\eta_{j|i}$. η is now used as an input. For $\eta \geq 0$, the diagonal elements of **M** are obtained by

$$m_{q,r} = \frac{1}{k} + \eta\left(1 - \frac{1}{k}\right) \quad , \quad q = r = 1, \ldots, k \qquad (16a)$$

and the off-diagonal elements by

$$m_{q,r} = \frac{1}{k-1}\left[1 - \left[\frac{1}{k} + \eta\left(1 - \frac{1}{k}\right)\right]\right], \quad q \neq r \qquad (16b)$$

For $\eta < 0$,

$$m_{q,r} = \frac{1}{k} + \eta\left(1 - \frac{1}{k}\right) \quad , \quad q = k - r \qquad (16c)$$

$$m_{q,r} = \frac{1}{k-1}\left[1 - \frac{1}{k} - \eta\left(1 - \frac{1}{k}\right)\right], \quad q \neq k - r \qquad (16d)$$

For instance, the link-strength parameter value 1 implies an identity matrix, 0 implies a non-informative link matrix, and 0.7 implies the following matrix, which is a 3 x 3 matrix for demonstration purposes:

$$\mathbf{M} = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .8 & .1 \\ .1 & .1 & .8 \end{bmatrix} \qquad (17)$$

Network Propagation. The algorithm for propagating uncertain information in the probabilistic layer (Varis 1994) is based on Pearl's (1988) polytree algorithm. Two independent messages (likelihoods) are computed, and the updated belief is obtained as the convolution product of these messages and the prior belief. The nodes are linked with link matrices that can be direction-specific. The polytree approach does not update messages in cases where the propagation direction is changed.

As was mentioned above, two information propagation directions can be distinguished in the network: top-down and bottom-up. The calculation is performed symmetrically, but directions up and down are used for verbal convenience.

When propagating messages downwards in a network, all messages coming to a node, say j, from an another node, say i, are denoted by $\mathbf{p}_{j|i}$ and messages leaving node i are denoted by π_i. For any node j, preconditioned by any node i ($i < j$):

$$\mathbf{p}_{j|i} = \mathbf{M}_{j|i}\, \pi_i \qquad (18)$$

The likelihood vectors $\mathbf{p}_{j|i}$ and π_i consist of the following elements:

$$\pi_i = \begin{bmatrix} \pi_i^1 \\ \pi_i^2 \\ \pi_i^3 \end{bmatrix} \quad \text{and} \quad \mathbf{p}_i = \begin{bmatrix} p_{j|i}^1 \\ p_{j|i}^2 \\ p_{j|i}^3 \end{bmatrix} \tag{19}$$

For elements r, the π_i^r message is the scaled vector product (joint distribution) of the message $\pi_i^r|_{1..i-1}$ and the evidence $\mathbf{e}_i^r$.

$$\pi_i^r = \pi_{i|1...i}^r = \alpha e_i^r \pi_{i|1...i-1}^r \tag{20}$$

where α is a scaling constant, scaling the sum of the k vector elements of π_i to unity. The incoming message $\pi_{i|1..i-1}$ is the joint distribution of all the messages, $\mathbf{p}_{i|1}$ to $\mathbf{p}_{i|i-1}$, from the node's $i - 1$ predecessors:

$$\pi_{i|1...i-1}^r = \prod_{k=1}^{i-1} p_{i|k}^r \tag{21}$$

Starting from the first node, the $\mathbf{p}_{1|0} = \mathbf{1}$ and $\pi_1 = \mathbf{e}_1$, $\mathbf{p}_{2|0,1} = \mathbf{M}_{2|1}\pi_1$, and so on.

Bottom-up propagation is quite similar to top-down propagation. Only the direction is reverse. All messages coming to node i from node j are denoted by $\mathbf{l}_{i|j}$ and messages leaving the node j are denoted by λ_j. For any node i, preconditioned by any node j, with $i < j$.

$$\mathbf{l}_{i|j} = \mathbf{M}_{i|j}\, \lambda_j \tag{22}$$

The λ_j message is the joint distribution of the message $\lambda_{j|j+1..n}$ and the evidence $\mathbf{e}_j$.

$$\lambda_j^r = \lambda_{j|j...n}^r = \beta e_j^r \lambda_{j|j+1...n}^r \tag{23}$$

where β is a scaling constant. The incoming message $\lambda_{j|j+1..n}$ is a convolution of all the messages, $\mathbf{l}_{j|j+1}$ to $\mathbf{l}_{j|n}$, from the node's $n - j$ successors:

$$\lambda_{j|j+1...n}^r = \prod_{k=j+1}^{n} l_{k|j}^r \tag{24}$$

For each node j, the posterior belief distributions $\mathbf{Bel}_j$ can now be calculated on the basis of the prior distribution $\mathbf{e}_j$, updating it with the information from the sub-network above and below the node, i.e., vectors $\pi_{j|1..j-1}$ and $\lambda_{j|j+1..n}$, respectively:

$$Bel_j^r = \gamma \pi_{j|1...j-1}^r e_j^r \lambda_{j|j+1...n}^r \tag{25a}$$

where γ is a scaling constant. The same equation can be written as a vector product of the two likelihood messages and the evidence vector:

$$\mathbf{Bel}_j = \gamma\, \pi_{j|1...j-1} \bullet \mathbf{e}_j \bullet \lambda_{j|j+1...n} \tag{25b}$$

3.4 Information from probabilistic layer to outcome layer

In the approach proposed, there are two different paths of information from the probabilistic layer to the outcome layer:

- The link strength parameter η is involved in the propagation of uncertainty (cf. Equation 12).
- The deviations between model prior distributions and posteriors $\mathbf{Bel}_i$ give important diagnostic information about the model. In parameter estimation or other adjustment of the model to fulfill given targets, the posteriors are iterated to make them uniform distributions.

A suggested quadratic/linear iteration scheme (Equation 26) providing rapid convergence is based on comparison of the probabilities of the different outcome values of a control variable. They are iterated to be equal to one another.

$$\mu_i^* = \mu_i + a \cdot \mathrm{cv}_i \cdot \left(Bel_y^1 - Bel_y^3\right)\left|Bel_y^1 - Bel_y^3\right| \qquad (26a)$$

$$\eta_i^* = \eta_i + b \cdot \left(Bel_y^2 - 1/k\right) \qquad (26b)$$

where a and b are convergence parameters, Bel^r_y is the posterior probability of outcome r, k is the number of outcomes, μ_i is the mean of the prior distribution of node i (a control variable), η_i is the estimated link strength, and $*$ refers to an updated iteration value. This iteration scheme was found to be essentially more rapid and practical than parametric approaches such as t-test based iteration.

4. An Example of Two-Directional Propagation

The purpose of this simple example is to demonstrate the two-directional uncertainty scheme used in belief networks. The model used contains only the probabilistic layer; the outcome layer and the use of decision variables have been left to other examples to ensure simplicity of this example.

The example comes from fisheries management. A rapidly increasing number of the world's commercially utilized fish stocks are under risk of being overexploited, due to growth in markets and improvement of equipment. Fish stock assessment is one of the major tasks in fisheries management, and is needed for reasonable fisheries restriction policies to safeguard the threatened stocks. Data collection from nature is most often out of the question due to high costs, and indirect data are typically used. This type of data tends to be corrupted by many types of biases. Decisions on allowable catches are needed regularly, often on an annual basis.

The simplest possible model for the system is one in which there are two mutually dependent variables: fish stock and fish catch per fishing unit (e.g., one fishing night; Figure 3). This dependency is usually used in assessment of both variables. There are several ways of obtaining independent information on the variables. In the present example, fish stock assessment is based on catch estimates and the number of returned taggings, and the catch assessment on stock estimates and taxation records of professional fishermen or enterprises. The outcomes of both variables are a 30% decrease from the previous year, unchanged level, and a 30% increase from previous year.

A methodologically interesting question arises from the fact that, in the scale under consideration, fish stock can be understood as the cause and fish catch as the effect. Assessment from cause to effect and vice versa is clearly a strength in any environmental and resource management task. In a longer time frame, over several years, there is also a feedback from fish catch to fish stock.

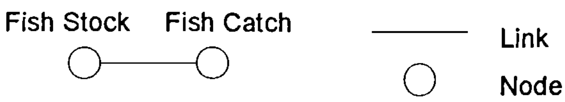


Figure 3. Structure of the example model.

The following notation is used: $\mathbf{e}_{stock}$ is the information from returned taggings, $\mathbf{e}_{catch}$ is the information from taxation records, π is the likelihood message from fish stock to fish catch, λ is the message from fish catch to fish stock, **M** is the link matrix which is equal in both directions, η is the corresponding link strength parameter, and α and β are scaling parameters. Now, we obtain the posteriors of the elements r of variables $\mathbf{Bel}_{stock}$ and $\mathbf{Bel}_{catch}$ by

$$Bel^r_{stock} = P(stock^r |\, e^r_{catch}) = \alpha\, P(stock^r)\, \lambda^r = \alpha\, e^r_{stock}\, \lambda^r \qquad (27a)$$
$$Bel^r_{catch} = P(catch^r |\, e^r_{stock}) = \beta\, P(catch^r)\, \pi^r = \beta\, e^r_{catch}\, \pi^r \qquad (27b)$$

The messages π and λ are

$$\pi = \mathbf{M}\, \mathbf{e}_{stock} \qquad (28a)$$
$$\lambda = \mathbf{M}\, \mathbf{e}_{catch} \qquad (28b)$$

Examine now the propagation of information in this model with several subsequent numerical cases. First, assume that the link strength is 0.6 and the information from returned taggings can be expressed as $\mathbf{e}_{stock} = [0.1,\ 0.3,\ 0.6]^T$, which tells us that the stock is likely to be increased. No other information is available (Figure 4). Using Equations 28 and 27, we obtain:

$$\pi = \mathbf{M}\,\mathbf{e}_{stock} = \begin{bmatrix} 0.73 & 0.13 & 0.13 \\ 0.13 & 0.73 & 0.13 \\ 0.13 & 0.13 & 0.73 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.3 \\ 0.6 \end{bmatrix} = \begin{bmatrix} 0.19 \\ 0.31 \\ 0.49 \end{bmatrix} \qquad (29a)$$

$$\lambda = \mathbf{M}\,\mathbf{e}_{catch} = \begin{bmatrix} 0.73 & 0.13 & 0.13 \\ 0.13 & 0.73 & 0.13 \\ 0.13 & 0.13 & 0.73 \end{bmatrix} \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} \qquad (29b)$$

$$\mathbf{Bel}_{stock} = \alpha\,\mathbf{e}_{stock} \bullet \lambda = 3 \begin{bmatrix} 0.1 \\ 0.3 \\ 0.6 \end{bmatrix} \bullet \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} = 3 \begin{bmatrix} 0.1 \cdot 0.33 \\ 0.3 \cdot 0.33 \\ 0.6 \cdot 0.33 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.3 \\ 0.6 \end{bmatrix} \qquad (29c)$$

$$\mathbf{Bel}_{catch} = \beta\,\mathbf{e}_{catch} \bullet \pi = 3 \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} \bullet \begin{bmatrix} 0.19 \\ 0.31 \\ 0.49 \end{bmatrix} = 3 \begin{bmatrix} 0.33 \cdot 0.19 \\ 0.33 \cdot 0.31 \\ 0.33 \cdot 0.49 \end{bmatrix} = \begin{bmatrix} 0.19 \\ 0.31 \\ 0.49 \end{bmatrix} \qquad (29d)$$

FISH STOCK (Cause)

y	e	Bel	λ
Decreased by 30%	0.1	**0.1**	0.33
Unchanged	0.3	**0.3**	0.33
Increased by 30%	0.6	**0.6**	0.33

M		
0.73	0.13	0.13
0.13	0.73	0.13
0.13	0.13	0.73

η 0.6

FISH CATCH (Effect)

π	Bel	e	y
0.19	**0.19**	0.33	*Decreased by 30%*
0.31	**0.31**	0.33	*Unchanged*
0.49	**0.49**	0.33	*Increased by 30%*

Figure 4. Propagation of fish stock information to fish catch.

Second, assume that we have information, instead of stock, on catch only. Now $\mathbf{e}_{catch} = [0.8,\ 0.15,\ 0.05]^T$. The propagation of this information proceeds as in the above case (Figure 5).

FISH STOCK

y	e	Bel	λ
Decreased by 30%	0.33	**0.61**	0.61
Unchanged	0.33	**0.22**	0.22
Increased by 30%	0.33	**0.16**	0.16

M		
0.73	0.13	0.13
0.13	0.73	0.13
0.13	0.13	0.73

η 0.6

FISH CATCH

π	Bel	e	y
0.33	**0.8**	0.8	*Decreased by 30%*
0.33	**0.15**	0.15	*Unchanged*
0.33	**0.05**	0.05	*Increased by 30%*

Figure 5. Propagation of fish catch information to fish stock.

Third, assume that we can use together the information that was available in the above cases separately. Figure 6 shows, that this controversial information forces both the belief vectors closer to non-informative ones than the values of the respective evidence vectors are.

FISH STOCK / FISH CATCH

y	e	Bel	λ	M			π	Bel	e	y
Decreased by 30%	0.1	**0.27**	0.61	0.73	0.13	0.13	0.19	**0.68**	0.8	*Decreased by 30%*
Unchanged	0.3	**0.3**	0.22	0.13	0.73	0.13	0.31	**0.21**	0.15	*Unchanged*
Increased by 30%	0.6	**0.43**	0.16	0.13	0.13	0.73	0.49	**0.11**	0.05	*Increased by 30%*
				η	0.6					

Figure 6. Impact of controversial information on posterior beliefs.

In the fourth case, the evidence vectors are no longer contradictory, but supporting one another. This results (Figure 7) a higher belief on increasing stocks and catches than the evidence vectors would alone indicate.

FISH STOCK / FISH CATCH

y	e	Bel	λ	M			π	Bel	e	y
Decreased by 30%	0.1	**0.07**	0.25	0.73	0.13	0.13	0.19	**0.1**	0.2	*Decreased by 30%*
Unchanged	0.3	**0.25**	0.31	0.13	0.73	0.13	0.31	**0.25**	0.3	*Unchanged*
Increased by 30%	0.6	**0.69**	0.43	0.13	0.13	0.73	0.49	**0.65**	0.5	*Increased by 30%*
				η	0.6					

Figure 7. Impact of mutually supporting information on posterior beliefs.

In the last two cases, the third case is revisited, but the link strength is changed, first to 0.4 and then to 0.8 (Figure 8). The former implies weaker association and the latter implies stronger association between the two variables, in comparison to the nominal case, in which the link strength is 0.6.

FISH STOCK / FISH CATCH

y	e	Bel	λ	M			π	Bel	e	y
Decreased by 30%	0.1	**0.2**	0.52	0.6	0.2	0.2	0.24	**0.73**	0.8	*Decreased by 30%*
Unchanged	0.3	**0.3**	0.26	0.2	0.6	0.2	0.32	**0.18**	0.15	*Unchanged*
Increased by 30%	0.6	**0.5**	0.22	0.2	0.2	0.6	0.44	**0.08**	0.05	*Increased by 30%*
				η	0.4					

FISH STOCK / FISH CATCH

y	e	Bel	λ	M			π	Bel	e	y
Decreased by 30%	0.1	**0.37**	0.71	0.87	0.07	0.07	0.15	**0.62**	0.8	*Decreased by 30%*
Unchanged	0.3	**0.29**	0.19	0.07	0.87	0.07	0.31	**0.24**	0.15	*Unchanged*
Increased by 30%	0.6	**0.34**	0.11	0.07	0.07	0.87	0.55	**0.14**	0.05	*Increased by 30%*
				η	0.8					

Figure 8. Impact of different link strengths on posterior beliefs.

The model was as simple as possible to demonstrate the two-directional propagation scheme in belief networks (probabilistic layer). In practice, fish stock assessment models are usually age-structured population models, which allow forecasting of stocks from one year to several years ahead. The basic problem setting, the simultaneous use of both stock and catch information, remains basically the same. An example of using belief networks, particularly the uncertainty balance approach, in fish stock assessment is given by Varis et al. (1993) for Baltic salmon.

5. An Example of Parameter Estimation by Uncertainty Balance

In the previous example, the model consisted only of two nodes in a probabilistic layer. In this example, we also include the outcome layer, targets (observations), and decision variables (parameters) in the analysis. To define the outcome layer, consider the following linear model

$$y_{i+1} = a\,y_i, \qquad i = 1, 2, 3 \tag{30}$$

where y_i is the model prediction of an observed variable e^*_i at point i, and a is a parameter. All these variables are normally distributed. The tasks are:

- To estimate the expected value of parameter a.
- To estimate the structural uncertainty of the model (link strengths).

These estimates are based on the three observations e^*_1, e^*_2, e^*_3. Figure 9 presents the structure of the model.

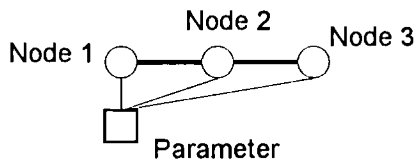


Figure 9. Structure of the example model.

5.1 *The model without external information*

In the following, the estimation procedure is illustrated with a numerical example, and the propagation scheme is calculated step-by-step. In the first step, a model is present with no observations. As it now includes the outcome layer and the probabilistic layer, it takes the form shown in Figure 10.

y_1	cv_{model}		y_2	cv_{model}		y_3	cv_{model}	
3	0.3	Model	3	0.3		3	0.3	
e^*_1	cv_{data}		e^*_2	cv_{data}		e^*_3	cv_{data}	
	0.3	Observation		0.3			0.3	

e_1	π_1	Bel_1	$l_{1\|2,3}$
0.33	0.33	**0.33**	0.11
0.33	0.33	**0.33**	0.11
0.33	0.33	**0.33**	0.11

$M_{1\|2} = M_{2\|1}$		
1	0	0
0	1	0
0	0	1
	$\eta_{1\|2} = \eta_{2\|1}$	1

e_2	$p_{2\|1}$	π_2	Bel_2	λ_2	$l_{2\|3}$
0.33	0.33	0.11	**0.33**	0.11	0.33
0.33	0.33	0.11	**0.33**	0.11	0.33
0.33	0.33	0.11	**0.33**	0.11	0.33

$M_{2\|3} = M_{3\|2}$		
1	0	0
0	1	0
0	0	1
	$\eta_{2\|3} = \eta_{3\|2}$	1

e_3	$p_{3\|1,2}$	Bel_3	λ_3
0.33	0.11	**0.33**	0.33
0.33	0.11	**0.33**	0.33
0.33	0.11	**0.33**	0.33

e_4	π_4	Bel_4	pl_4	$pl_{4\|1}$	$pl_{4\|2}$	$pl_{4\|3}$
0.33	0.33	**0.33**	0.33	0.33	0.33	0.33
0.33	0.33	**0.33**	0.33	0.33	0.33	0.33
0.33	0.33	**0.33**	0.33	0.33	0.33	0.33

a	cv_{model}	
1	0.3	Parameter

$M_{4,123}$		
1	0	0
0	1	0
0	0	1
	$\eta_{4\|1,2,3}$	1

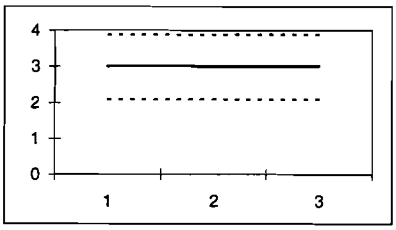


Figure 10. The model with no external information.

Due to the discrete approximation principle described earlier (Figure 2), all the distributions are uniform by definition if no external information is included in the analysis. Due to this definition, changes introduced in parameter values or initial states of the model do not introduce changes in the probabilistic layer. Changes in link strength values change only values in the link matrices, but do not influence any of the probability distributions.

5.2 *One observation is included*

When adding an observation at any node — say, node 1 as an example — the continuous distribution of the observation e^*_1 is approximated with a discrete distribution having the same outcome values as were used in the discrete approximation of the model output distribution at node 1 (Equations 13-15).

The information in $\mathbf{e}_1$ is included in the π message, and is now propagated through the network (Figure 11). Note that the posterior distributions (**Bels**) now equal the π messages, because there is no information coming up to the λ system. The non-uniform distributions in the probabilistic layer imply that there is also other information available besides the model, the posterior of the parameter $\mathbf{Bel}_4$ is also non-uniform. This feature will be used later in parameter estimation.

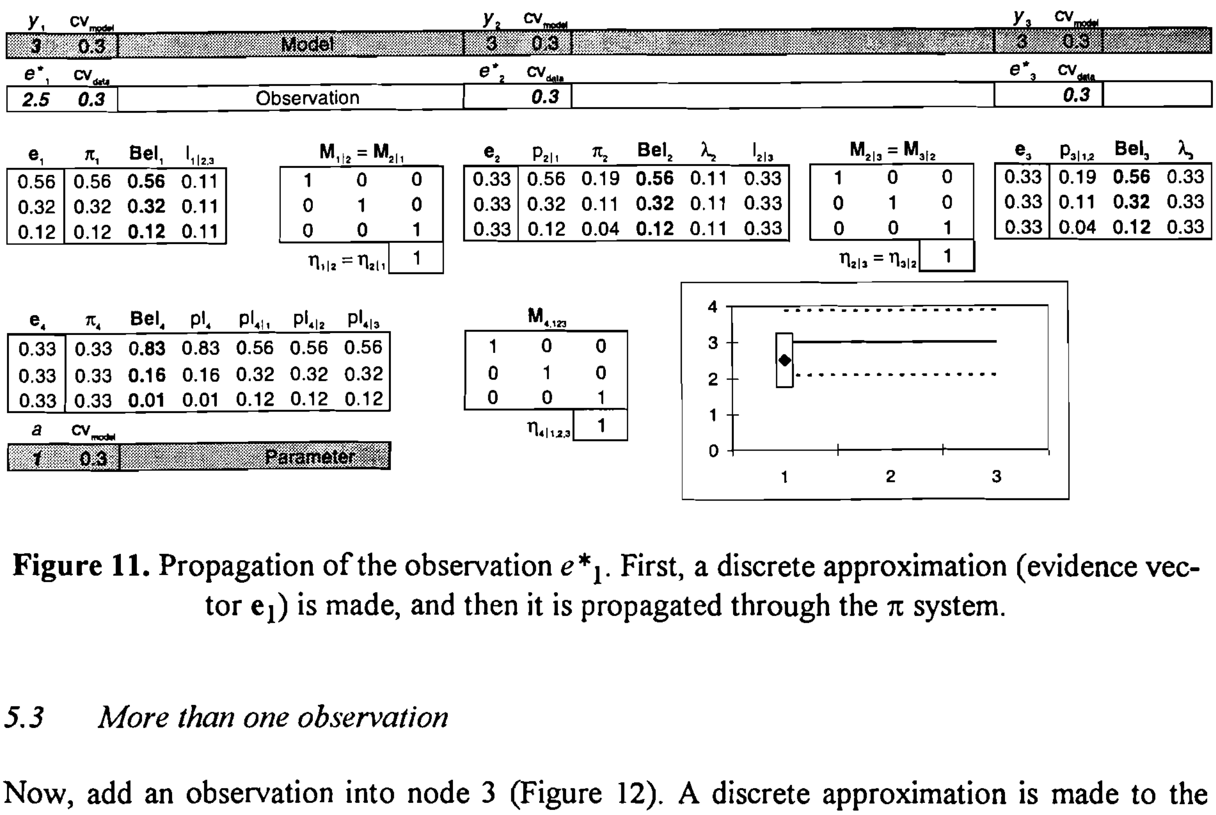


Figure 11. Propagation of the observation e^*_1. First, a discrete approximation (evidence vector $\mathbf{e}_1$) is made, and then it is propagated through the π system.

5.3 *More than one observation*

Now, add an observation into node 3 (Figure 12). A discrete approximation is made to the distribution of e^*_3, and the information is propagated through the network. Correspondingly, we can add an observation to node 2 (Figure 13). Note that the **Bels** are no longer equal to either the π or the λ messages, but their scaled vector product. The posterior of parameter $\mathbf{Bel}_4$ is again updated.

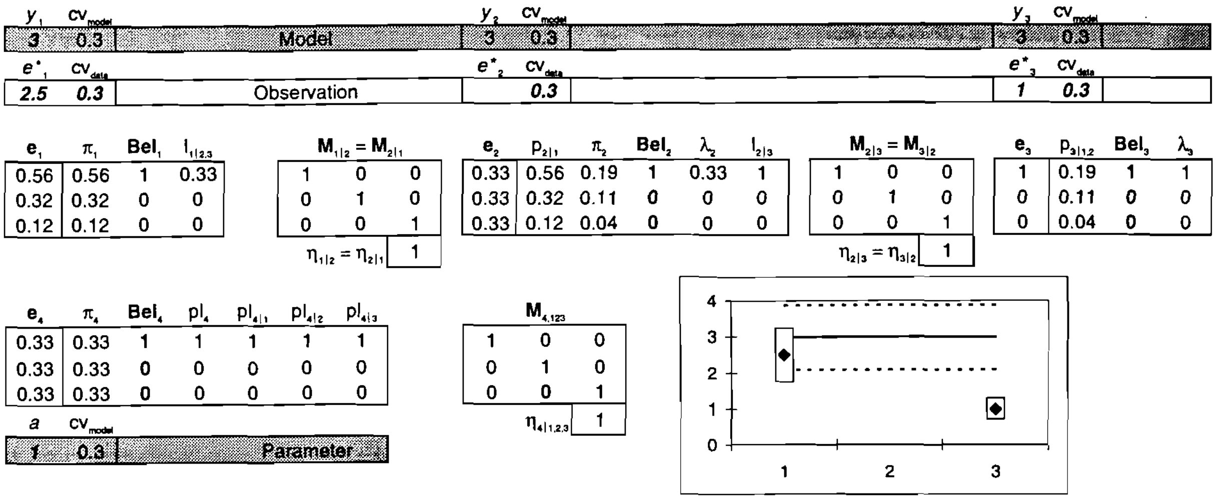


Figure 12. Propagation of the observation e^*_3. First, a discrete approximation (evidence vector $\mathbf{e}_3$) is made, and then it is propagated through the λ system.

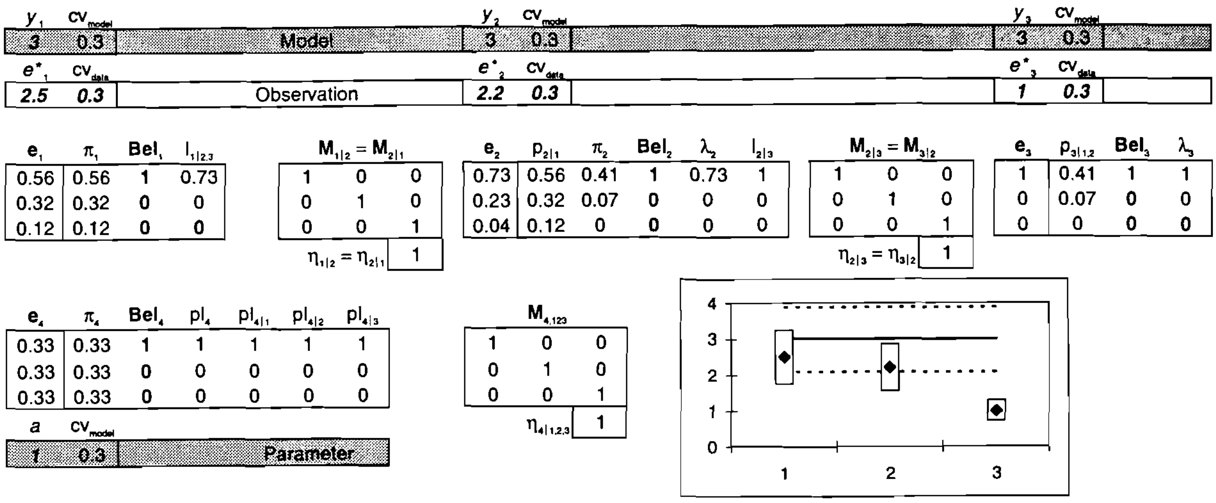


Figure 13. Propagation of the observation e^*_2. First, a discrete approximation (evidence vector $\mathbf{e}_2$) is made, and then it is propagated through the π system to the direction of node 1, and through the λ system to the direction of node 3.

5.4 *Parameter estimation*

This step estimates (1) a value to the parameter and (2) the link strengths between nodes 1 and 2, and nodes 2 and 3 (Figure 9). The principle used can also be applied to many other optimization tasks, as will be shown in the river water quality example later on. The idea is to obtain such values to the parameter and the link strength that $\mathbf{Bel}_4$ becomes uniform. In Figure 14, this iteration has been done. Figure 15 gives a set of examples of possible distributions of $\mathbf{Bel}_4$, and of the inference that can be made on the basis of such distributions. Note that when either a parameter value, link strength value, or observed value is changed, the probability values in the evidence vectors are also changed, because the model outcome distributions change. Evidently, the initial value of the model can also be defined as a control variable (parameter). Its value would then be iterated in a way similar to any other parameter (Equation 26).

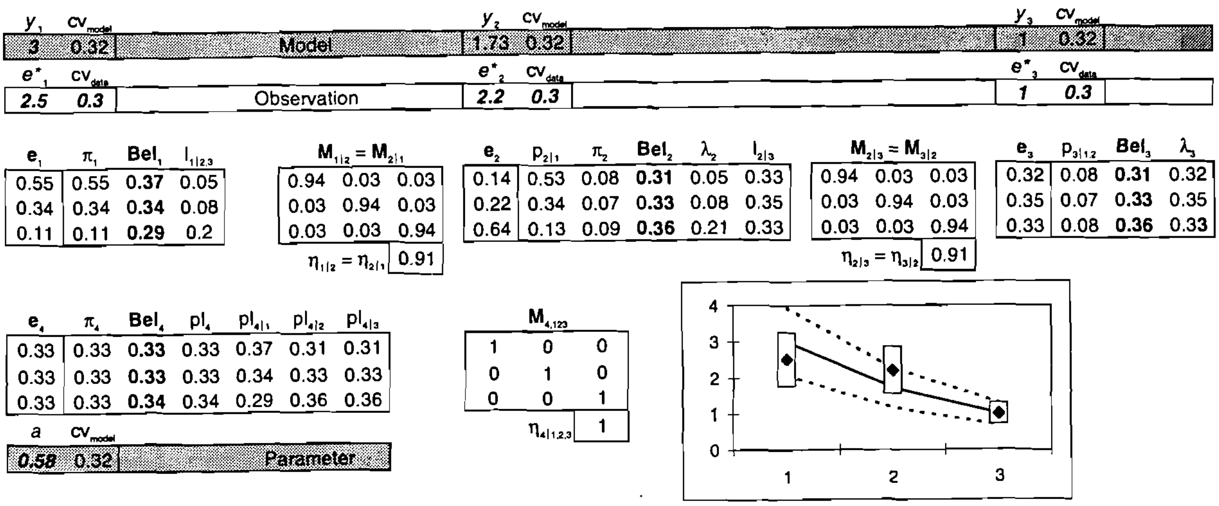


Figure 14. The model after iteration of $\mathbf{Bel}_4$ to be a uniform distribution.

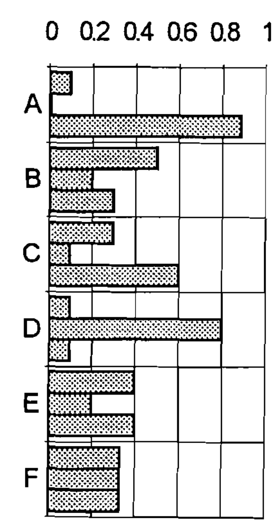
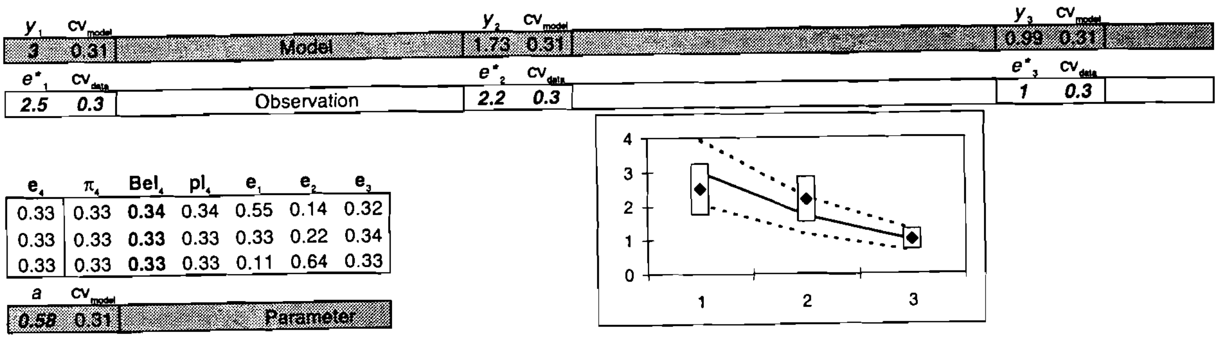
	E(parameter)	*Link strength*
 A	too low	too high
B	too high	too high
C	too low	too high
D	OK	too low
E	OK	too high
F	OK	OK

Figure 15. Some example posterior distributions of the parameter ($\mathbf{Bel}_4$) and the inference based on these types of distributions.

5.5 *Simplifying the configuration of the probabilistic layer*

Clearly, the configuration of a belief network relates to the use of the model. In general, a network offers a very flexible way of defining nodes to decision variables, but the same nodes do not need to be decision variables throughout the analysis. An illustration of this feature is given in the river water quality example below. First, parameters and link strengths are estimated, i.e., they are the decision variables. Thereafter, wastewater treatment levels at various treatment plants were used as decision variables.

The above example can be simplified considerably, if we have no any other purpose for the modeling task but model calibration. The same parameter and link strength values as above, with high accuracy, can be obtained directly as the joint distribution of evidence vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ (Figure 16). This configuration of the net is much less computation intensive and is thus remarkably faster in updating, yet it offers remarkably fewer options and possibilities for further studies with the model.



$\mathbf{e}_4$	π_4	$\mathbf{Bel}_4$	$\mathbf{pl}_4$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
0.33	0.33	**0.34**	0.34	0.55	0.14	0.32
0.33	0.33	**0.33**	0.33	0.33	0.22	0.34
0.33	0.33	**0.33**	0.33	0.11	0.64	0.33

Figure 16. A simplified configuration of the probabilistic layer. $\mathbf{Bel}_4$ is calculated as a vector product of evidence vectors $\mathbf{e}_1$ to $\mathbf{e}_3$.

6. An Example of River Water Quality Management

The previous examples illustrated the different propagation features, the two-layers concept, and the uncertainty balance iteration principle. The third example is more comprehensive, and it has been constructed to correspond to a real-life resource management problem.

6.1 The management problem and the watershed

The example deals with cost-effective upgrading of wastewater treatment plants in a watershed on the basis of ambient water quality criteria. It is intended to represent a typical river basin management problem, particularly in conditions of Central and Eastern Europe. The formerly socialist countries of the region are in the midst of a very rapid and profound transition process, affecting almost all conceivable sectors of the society. Clearly, this also includes water quality management (Somlyódy et al. 1994). Previously, the integration of ambient and effluent monitoring has been low. At present, the industry is undergoing considerable change, and past water quality data are of limited validity, yet there is a pressing need for improving water pollution control. The scarcity of capital suggests the policy of gradual upgrading of wastewater treatment on a cost-effective basis (Somlyódy 1993).

A hypothetical watershed is used with ten municipal wastewater treatment plants which should be upgraded to improve the river water quality. Each plant discharges the effluent into a different tributary. The impact of different ambient water quality criteria and diverse investment levels should be studied under the precepts of cost-effective prioritization of upgrading levels at different plants. The hypothetical data are presented in Appendix 2. A variety of treatment alternatives is available (Tables A3-A5), ranging from no treatment (0) and biological treatment (1) to more advanced solutions (cf. Somlyódy et al. 1994). Initially, all the plants are at level 1.

6.2 A probabilistic river model

Based on the results of the comprehensive water quality management study of the Nitra River Basin, Slovakia (Masliev and Somlyódy 1994, Somlyódy et al. 1994), an extended Streeter-Phelps model with three state variables was chosen for this study. The state variables are dissolved oxygen (DO), and carbonaceous and nitrogenous biological oxygen demand (BOD and NH4, respectively). Three parameters are estimated, including BOD oxygenation rate, reaeration coefficient, and NH4 oxygenation rate, which are not stretch-specific. The state equations of the model constitute the outcome layer of the system. When calculating the total costs of upgrading the treatment plants, an interest rate of 6% and an economic life of 20 years for the project were assumed. A standard capital recovery rate factor was used when transforming investment costs to annual costs.

The probabilistic layer is based on a series of parallel, coupled probability trees based on the river topology (Figure 17), which describe the steady-state evolution of the state variables. State variables and parameters are represented as belief network nodes. Evidential information for states is obtained from data (Appendix 2).

The analysis is divided into two subsequent phases, at both of which the uncertainty balance iteration approach is used. The same model including the two layers is used, but the targets, decision variables, and estimated link strengths are different (Table 1, Figures 18 and 19). First, the parameter estimation is performed, in which the mean values at the outcome layer are iterated to equal the posteriors. The link strengths of the links shown in Figure 18 are estimated; their values show the structural uncertainties of the state equations.

The second phase consists of finding the most cost-effective solutions for river water quality management, taking into account the water quality targets for the river and the costs involved. Now, different treatment levels are used as decision variables (instead of parameters at the previous phase), link strengths are not estimated, and water quality targets together with the target cost level are used as targets (vs. observations at the previous phase).

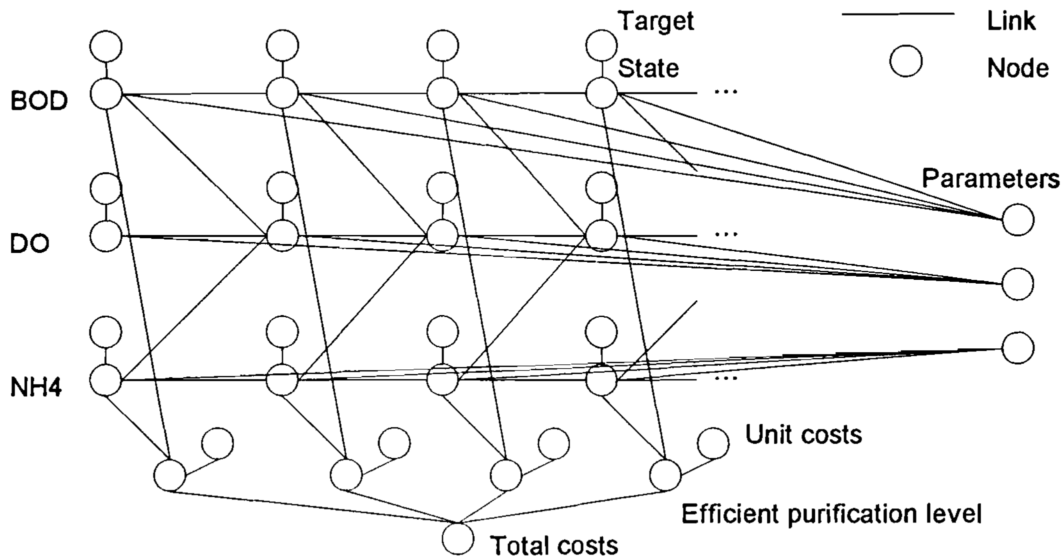


Figure 17. Configuration of the probabilistic layer.

The definition of variables can be changed in the course of the analysis due to the two-directional uncertainty propagation in the probabilistic layer. At the diagnostic phase, both downstream and upstream (π and λ, respectively) messages are used (in the same manner as in the second example; Figures 10-12). In the management support phase, only the λ message going upstream is used. A clear logical explanation exists on the use of these propagation principles. In the diagnosis, all the data and model predictions are iterated to meet a balance, hence both propagation directions are used. In the management support phase, the targets influence only the treatment plants downstream of the point at which a target is set. When detecting a deviation between target and model prediction, the message induced is propagated upstream all the way to the posterior distributions of the treatment plant purification levels. This provides a basis for iteration similar to that in parameter estimation.

Table 1. Definition of decision (control) variables and targets in the diagnostic and in the management support parts of the study.

	Diagnosis	*Management model*
Decision variables	Parameters	Dischargers
Targets	Observed water quality	Water quality targets
		Target Costs

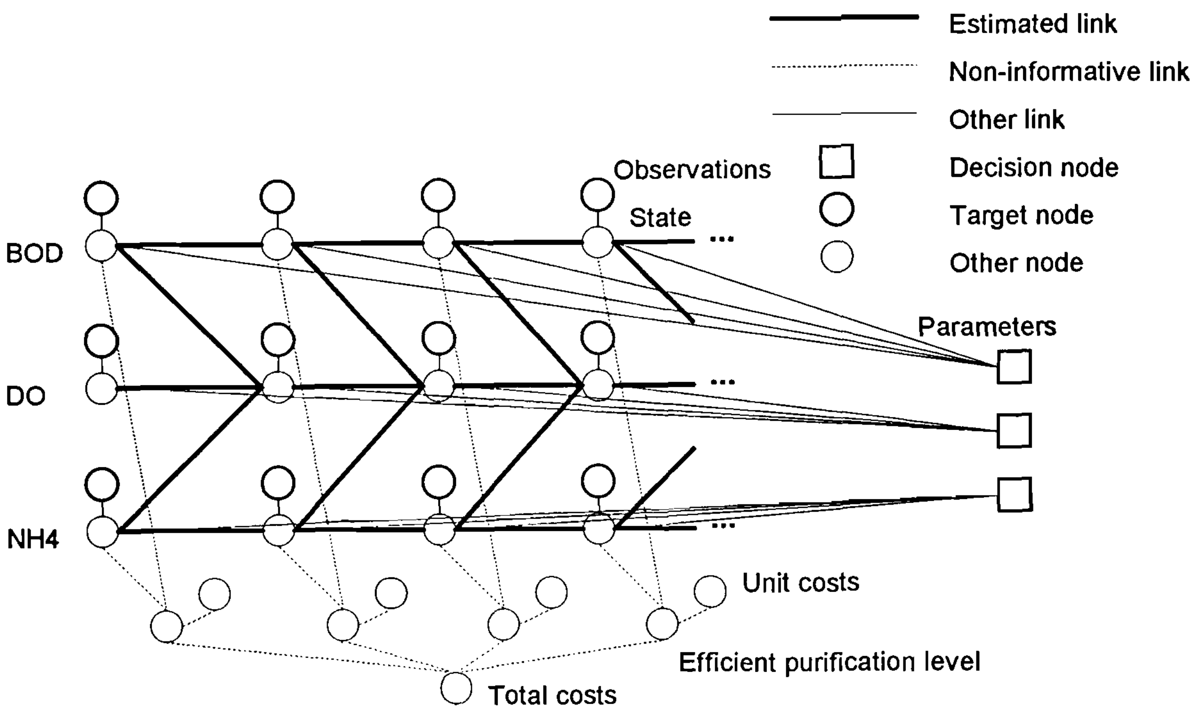


Figure 18. Probabilistic layer in diagnostic phase. The targets are now the observations, and the control variables are the model parameters.

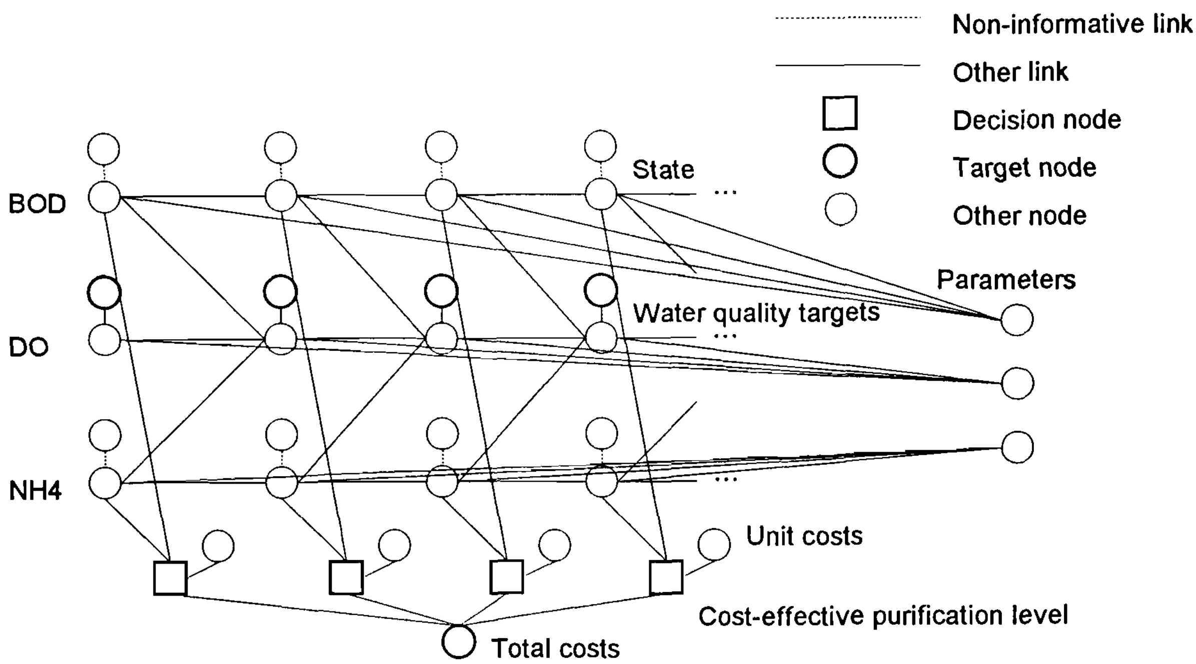
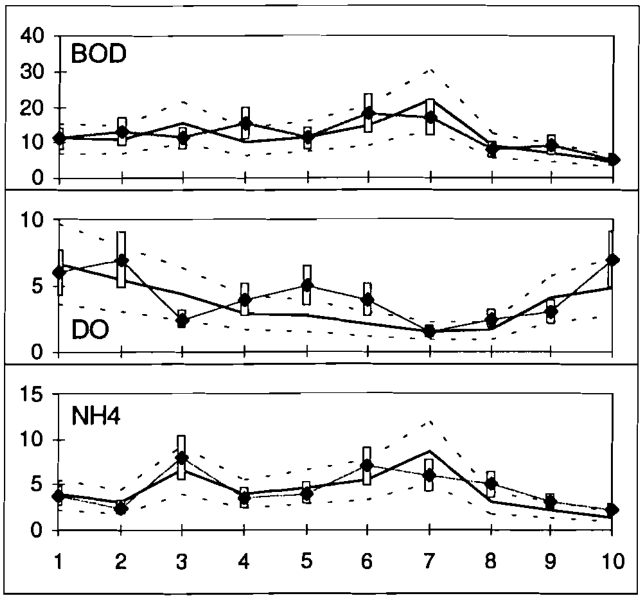


Figure 19. Probabilistic layer in management support phase. The targets are now the ambient water quality criteria and the total costs, and the control variables are the purification levels.

6.3 Model vs. data: illustration of the approach

The hypothetical data in Appendix 2 define our nominal case. For illustrative reasons, it was constructed so as not to excessively contain any features discussed below. The model calibration for the hypothetical data is shown in Figure 20.



	$E(k_i)$	Links
BOD	0.209	0.826
DO	0.414	0.869
NH4	0.258	0.807

Figure 20. The nominal case.

We take an example of the propagation of evidential information (observations) in one of the model equations, say BOD (Figure 21). We are interested in knowing how well the model is modeling our system and we want the model/system correspondence to be as good as possible. The system reference consists, in this case, of observations. In case *A*, there are no observations, and the probabilistic layer consists of uniform distributions, and we have no information on the success in our modeling task. In case *B*, there is one observation available, and its probability distribution is discretized according to the procedure described in subsection 3.2 (cf. Figure 2., Equations 13-15). This vector is used as an evidence vector, and this information is propagated throughout the net as Figure 21 shows. Case *C* includes one more observation, which is again discretized and fed into the net. Case *D* includes a series of observations. In cases *B*, *C*, and *D*, the parameter estimation is based on the iteration of the joint distribution (scaled vector product) of the **Bels** to a uniform distribution. The more differences there are in the column heights on the right-hand figures, the more misfit there is between model and data. Clearly, the better the model fit, the lower is its structural uncertainty.

We next demonstrate the influence of prior uncertainties and model fit on posterior uncertainties (Table 2, Figure 22). The data may have a low uncertainty, but the model fit is poor. In such a case (case *A*), the link strengths become low, structural uncertainty of the model becomes high, and the model prediction highly uncertain. If the data have high uncertainty and the fit is poor, the link strengths may still be high, but model prediction remains uncertain (case *B*). If the data have low uncertainty and the model fit is good, then the link strengths are high and model prediction has low uncertainty (case *C*). If the data have high uncertainty and the model fit is good, then the link strengths are high and model prediction has high uncertainty (case *D*)

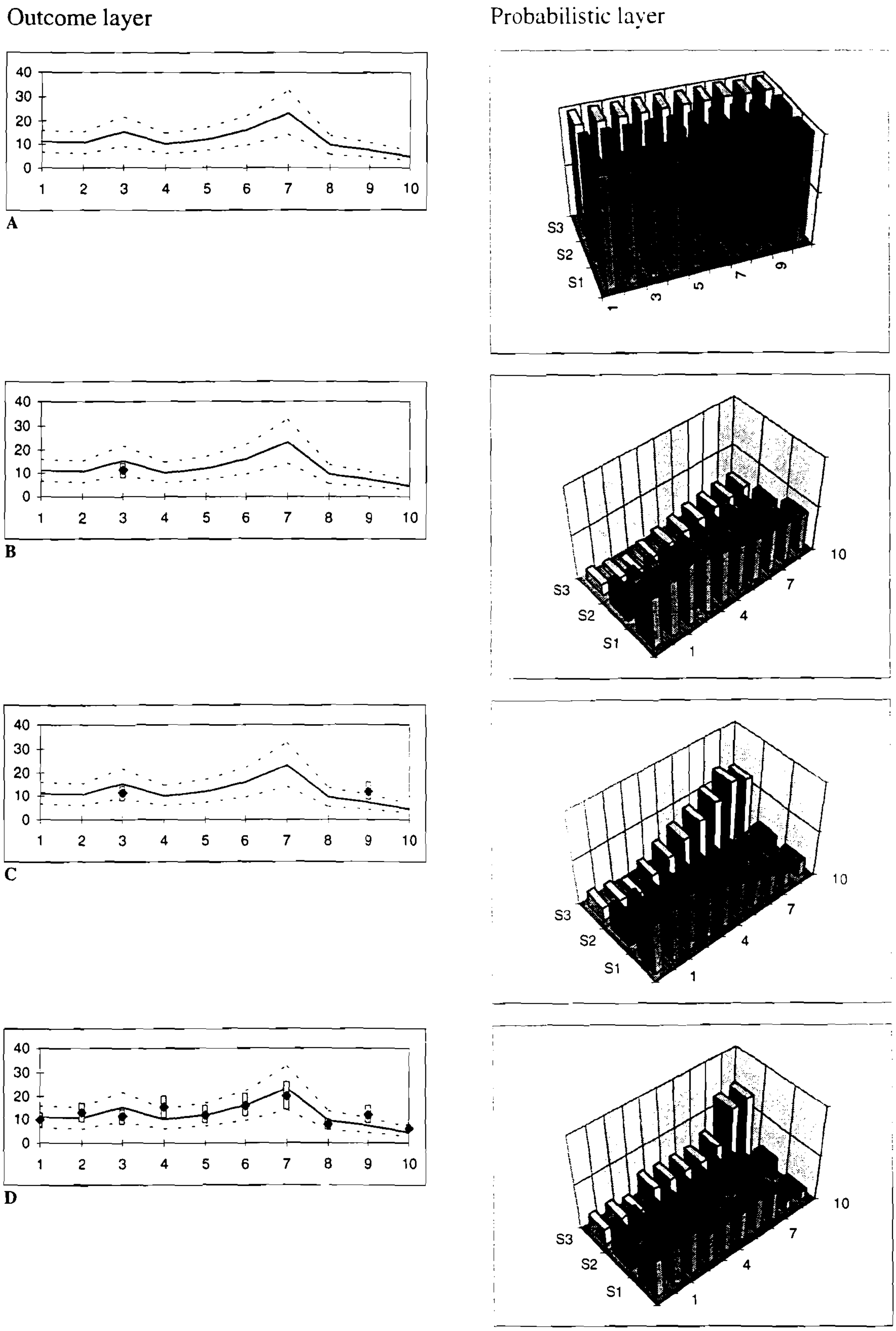
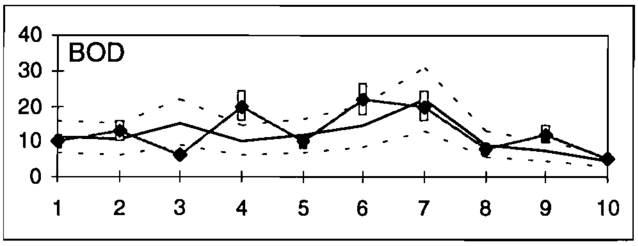


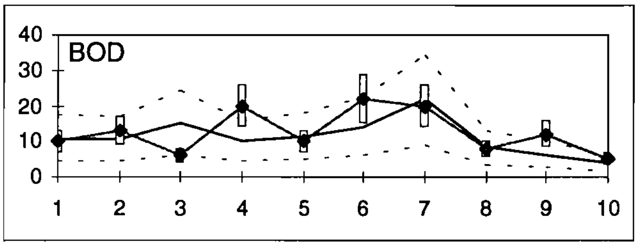
Figure 21. Propagation of observed information in the probabilistic layer. Each observation is discretized to be an evidence vector which is then fed into the probabilistic layer. The columns on the right show the posterior distributions (**Bels**) of the model prediction at different points.

Table 2. Different typical combinations of prior information and their influence on posterior information (cf. Figure 22).

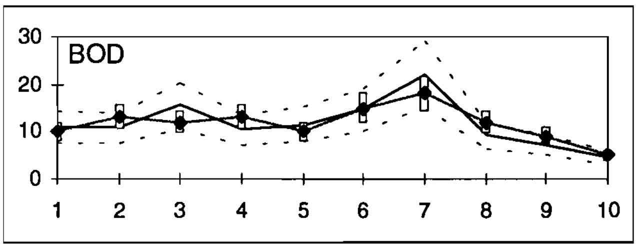
Case	*Priors*		*Posteriors*	
	E(data) vs. E(model prediction)	*Uncertainty(data)*	*link strengths*	*Uncertainty (model prediction)*
A	low accordance	low	low	high
B	low accordance	high	high	high
C	high accordance	low	high	low
D	high accordance	high	high	high



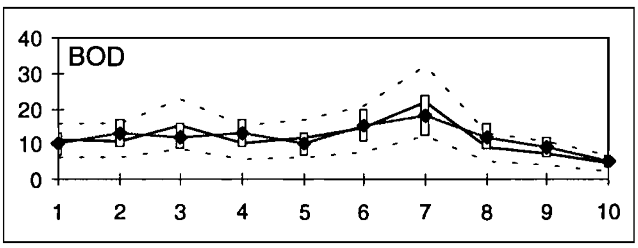
A: cv(BOD) = 0.28, Links = 0.36



B: cv(BOD) = 0.42, Links = 0.68



C: cv(BOD) = 0.28, Links = 0.81



D: cv(BOD) = 0.42, Links = 0.86

Figure 22. Example with the BOD equation: The influence of uncertainties and controversies in prior information on uncertainties in the posterior information (cf. Table 2).

The issue becomes more complex when more than one state equation is included in the model, as is the case with our river model. Here, the state equations are inter-linked so that the parameters for the BOD and NH4 equations must be estimated first, and thereafter the DO equation is in turn. Success in the BOD prediction is strongly dependent on success in predicting the BOD and NH4 concentrations along the river. However, it may often happen that the uncertainties within the BOD and NH4 data are much higher than those within the DO data, due purely to analytical accuracy and variability of different substances in nature. Under such conditions, the uncertainties of the BOD and NH4 predictions become high and are propagated throughout the state equations (outcome layer) to the DO prediction, which also becomes highly uncertain (Figure 23). This occurs despite good empirical evidence. In such a case, our model structure might no longer be as efficient and another model configuration could be considered, for instance, an empirical model based more strongly on high-quality empirical evidence.

Often, for different reasons, a prefixed set of parameter values is used in the analysis. These values might be used in standard fashion or when the use of literature values would be considered more adequate than that of empirical parameters. In such a case, the link strength becomes lower than it would be if empirical parameter values were used, unless they were equal. Accordingly, we pay a price for using standard parameter values under conditions of higher uncertainty in prediction.

At the management analysis phase, there can also be analogous cases as above. The largest difference in our example is, however, that link strengths are not estimated at this phase. An interesting phenomenon occurs if the target economic level is set too high compared with the ambient water quality targets. The approach does not find a single solution, because there is looseness in the targets. Either the economic targets should be set lower or the ambient targets should be higher, or both should be done to find a single solution.

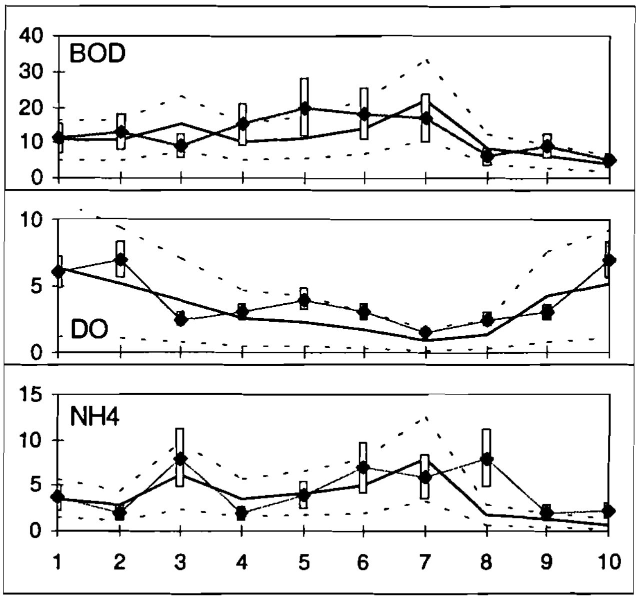


Figure 23. The DO prediction remains highly uncertain even though DO observations are very accurate, but if BOD and/or NH4 predictions have high uncertainty.

6.4 River water quality management scenarios

The model is now used to produce scenarios with diverse targets for ambient water quality and varying economic targets. Support for prioritization of the upgrading of wastewater treatment plants is required.

There are many ways of setting ambient water quality targets. The widely applied principle of equity is usually understood as an equal purification level for all plants, for instance, as equal effluent quality criteria, disregarding ambient standards. Equity can also be understood as a requirement for improving the water quality equally in all parts of the river, although this is not a common interpretation. An often used and perhaps the most typical concept is to set a minimum target level for DO. One can also set site-specific targets, for instance, at a location important as recreational site or water intake area. With a probabilistic model, levels which are not the expected value scenarios can also be considered. It may sometimes be more important to strive to keep DO concentrations above a certain level in 90% vs. 50% of the critical cases. In summation, Figures 24-28 show examples of five sets of scenarios:

- Equal purification level (equity: dischargers).
- Equal improvement (equity: river).
- Minimum concentration level.
- Target(s) at specific point(s) (city, outflow, water intake, recreational area, protected site, etc.).
- Target probability levels (frequencies of occurrence).

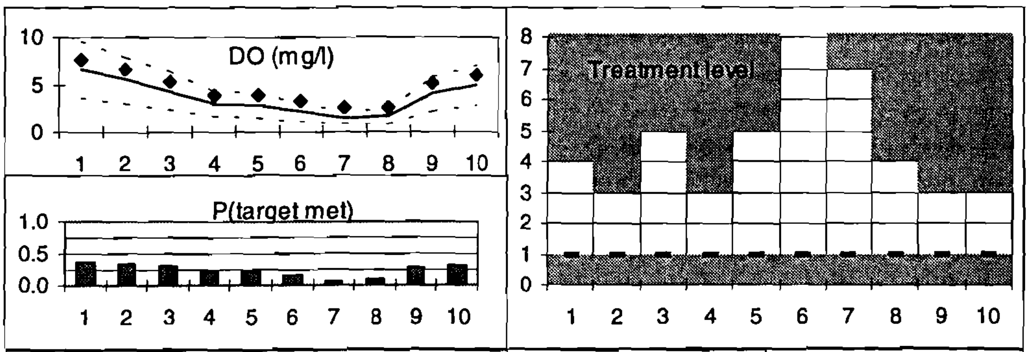
In these scenarios, five different target levels of costs were used (Table 3).

Table 3. Target levels for total costs due to upgrading of the wastewater treatment plants.

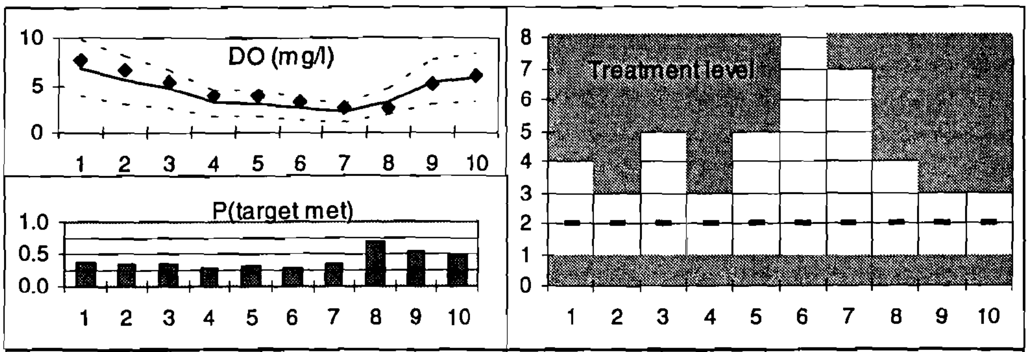
Cost range	*Equal cost equivalent treatment level*
5	initial, all 1
6.5-6.7	all 2
8.5-8.7	intermediate between 2 and 3
9.3-9.8	all 3
15.6	all at maximum level

The most important feature seen in these scenarios is that normative, equal treatment level at all plants is the most costly way of improving ambient water quality among the studied options, which clearly justifies the problem setting. Another issue is whether the prioritization is workable institutionally and juridically in the countries of the region.

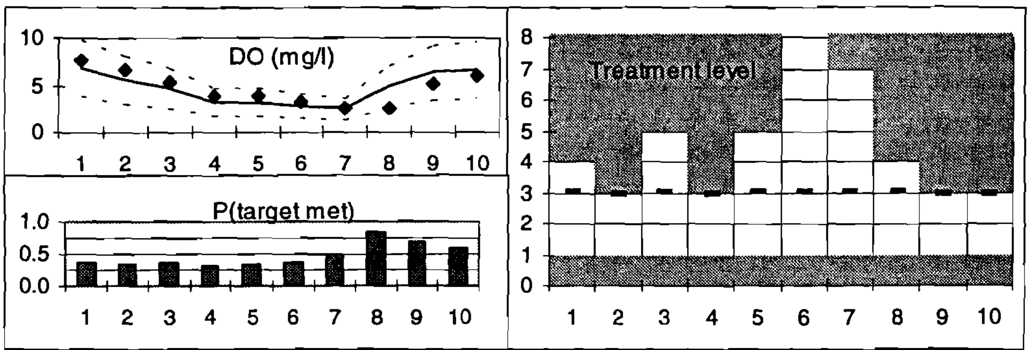
Despite the rather divergent results of the different scenarios, there appear to be certain plants that tend to gain high priority and also those that systematically have low priority. The former feature is shared by Plants 6, 3, and 7, and to a certain extent also by Plant 1. The study would suggest prioritization of these plants compared with the others. A rather interesting observation is that despite a rich variety of different ambient water quality criteria and a wide range of target cost levels, the results show that there is no cost-effective basis for upgrading Plants 4, 9, and 10 in any case. Despite the fact that setting ambient water quality targets is often rather ambiguous, a promising result as far as water treatment prioritization is concerned, is that the results show certain common features regardless of targets used.



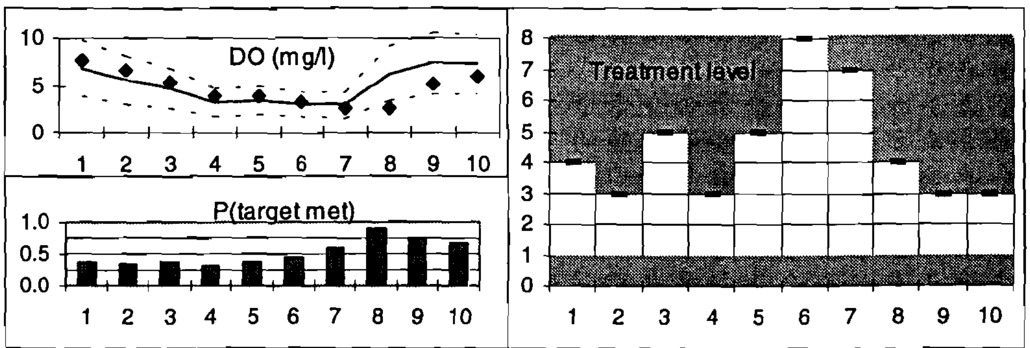
Scenario	Actual costs	Ambient targets	min [P(target met)]
Present state	4.9	Improve 1 mg/l	0.07



Scenario	Actual costs	Ambient targets	min [P(target met)]
Upgrade all to 2	6.6	Improve 1 mg/l	0.29

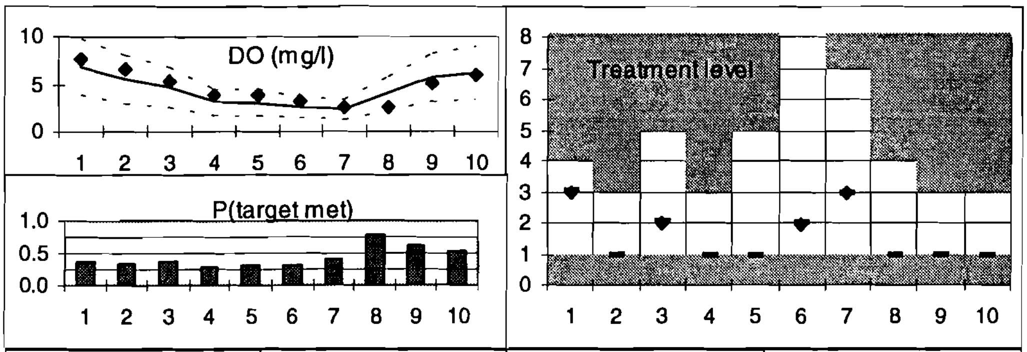


Scenario	Actual costs	Ambient targets	min [P(target met)]
Upgrade all to 3	9.3	Improve 1 mg/l	0.31

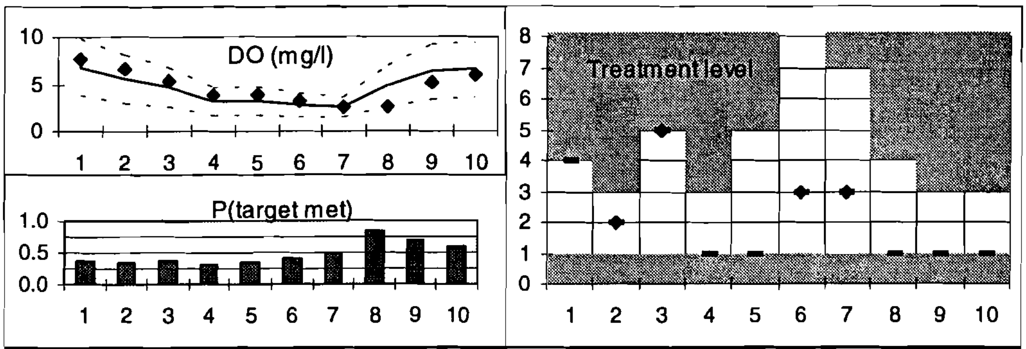


Scenario	Actual costs	Ambient targets	min [P(target met)]
Upgrade all to MAX	15.6	Improve 1 mg/l	0.32

Figure 24. Example scenarios on equal treatment level at all plants, and scenario for maximal treatment level at all plants.

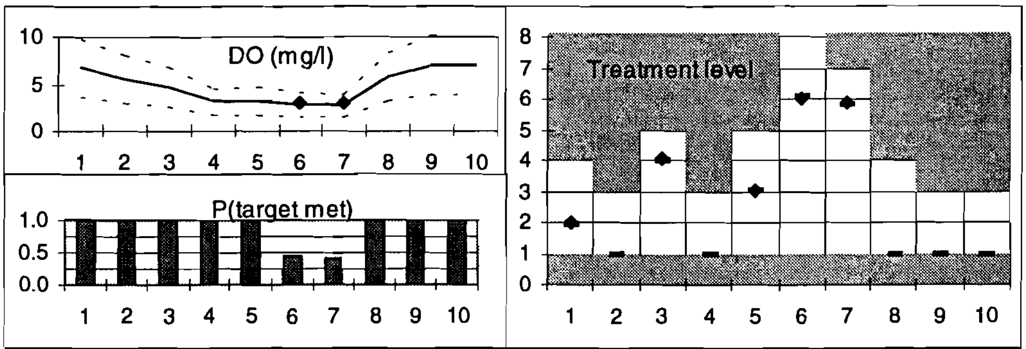


Scenario	Actual costs	Ambient targets	min [P(target met)]
Equity in river 1	6.5	Improve 1 mg/l	0.29

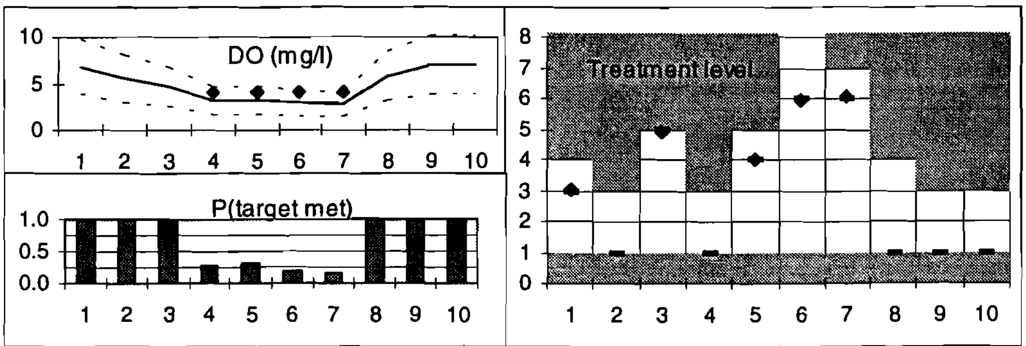


Scenario	Actual costs	Ambient targets	min [P(target met)]
Equity in river 2	8.6	Improve 1 mg/l	0.32

Figure 25. Example scenarios on equal improvement level at all parts of the river.

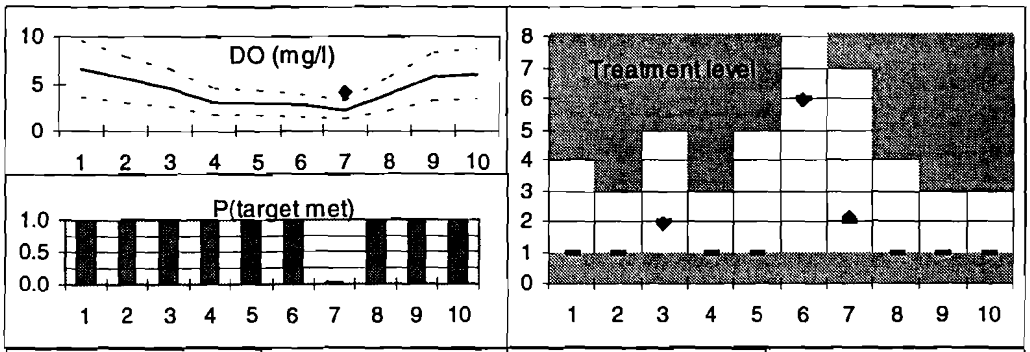


Scenario	Actual costs	Ambient targets	min [P(target met)]
Minimum level 1	8.7	min 3 mg/l	0.41

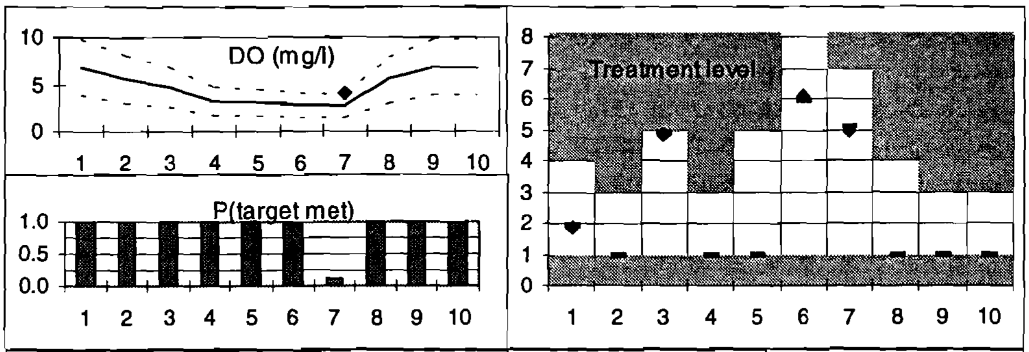


Scenario	Actual costs	Ambient targets	min [P(target met)]
Minimum level 2	9.8	min 4 mg/l	0.17

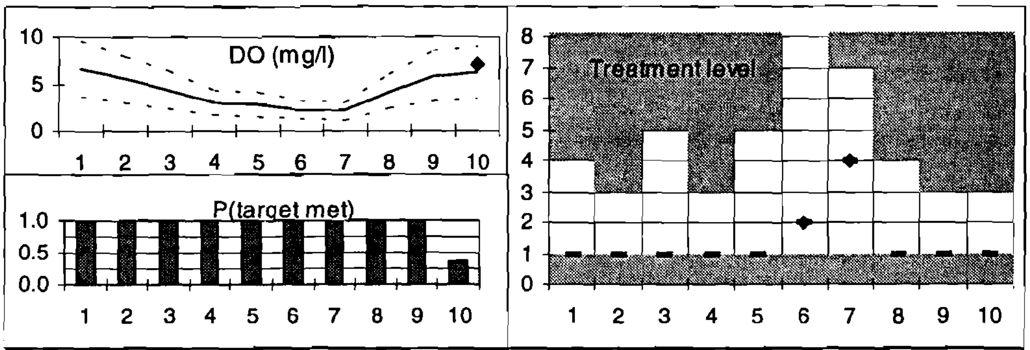
Figure 26. Example scenarios on minimum concentration level over the whole river.



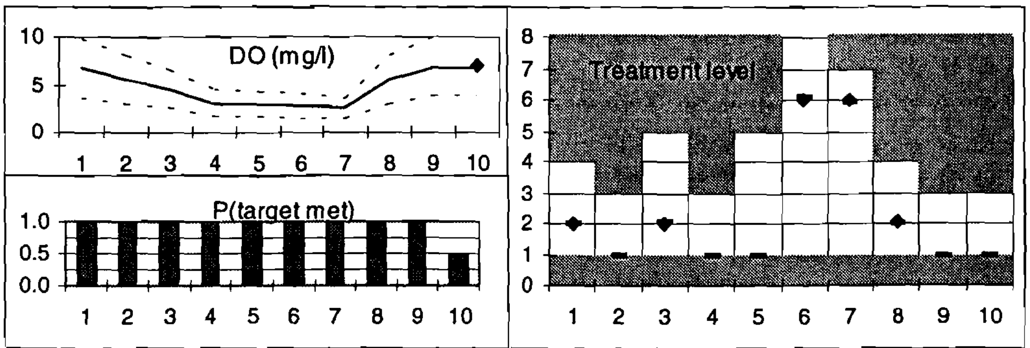
Scenario	Actual costs	Ambient targets	min [P(target met)]
Site specific 1	6.5	min 4 mg/l at point 7	0.03



Scenario	Actual costs	Ambient targets	min [P(target met)]
Site specific 2	8.6	min 4 mg/l at point 7	0.13

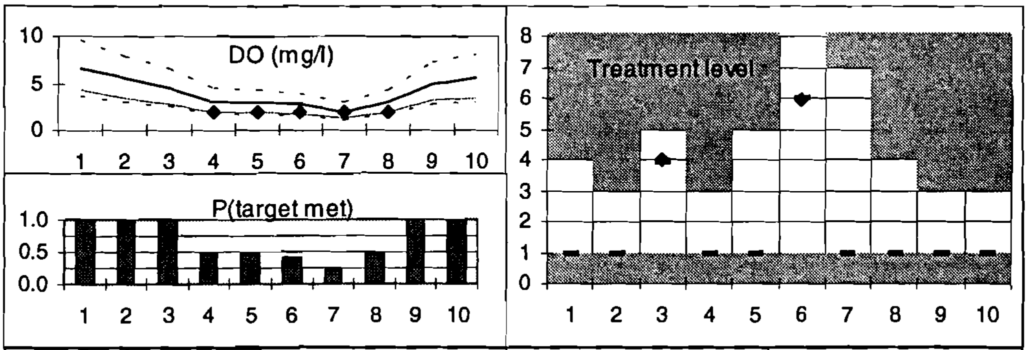


Scenario	Actual costs	Ambient targets	min [P(target met)]
Site specific 3	6.7	min 7 mg/l at point 10	0.38

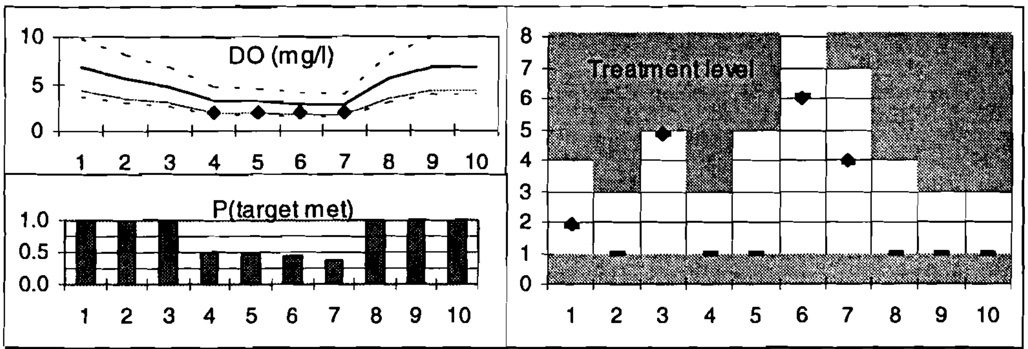


Scenario	Actual costs	Ambient targets	min [P(target met)]
Site specific 4	8.6	min 7 mg/l at point 10	0.48

Figure 27. Example scenarios on site specific ambient water quality targets.



Scenario	Actual costs	Ambient targets	min [P(target met)]
Risk averse, 80% 1	6.5	min 3 mg/l at river	0.21



Scenario	Actual costs	Ambient targets	min [P(target met)]
Risk averse, 80% 2	8.7	min 3 mg/l at river	0.39

Figure 28. Example scenarios on the use of different risk levels in ambient water quality targets.

7. Discussion and Conclusions

An approach to using belief networks for probabilistic modeling in resource and environmental management is presented. The approach enables updating of uncertainties in different model components interactively. It is therefore remarkably rapid, particularly when compared with conventional approaches requiring off-line simulation runs. Such approaches as Monte Carlo or Latin Hypercube simulations, although practicable in many cases, are time-consuming, non interactive, and have been criticized as being rather inaccurate (e.g., Morgan & Henrion 1990). The major disadvantage of the proposed approach is the relatively labor-intensive computer implementation when compared with conventional simulation approaches, at least at the pilot-study phase documented here.

The proposed approach can be used to detect inconsistencies among different pieces of information in different model components. Possible inconsistency appears as a difference between Bayesian prior and posterior distributions, in a given model component. This feature was used to develop an optimization approach in which prior and posterior distributions of objective functions are iterated to become equal. This can be done by changing the values of control variables and adjusting linking properties in the belief network.

The uncertainty balance approach can handle more than one objective function simultaneously. For instance, in the river basin example, the management optimization included two objectives: target costs and target ambient water quality. The approach finds a compromise (trade-off) between these targets, and can thus be used as a multiobjective optimization approach.

Within environmental and resource management sectors, common practical management models are relatively simple constructs that often can be analytically solved. Apparently, the use of relatively simple and well-known or easily comprehendable, conceptual models is a great advantage in practical assessment and management modeling. Transparency and quality assurance are often critical points when striving for the proper, critical attitude and utilization of modeling results. The proposed approach allows consideration of such models as uncertain constructs. The structural uncertainty can be estimated empirically, and the models can be linked and fused with other pieces of probabilistic information.

In the management of natural resources and the environment, the uncertainties are often very high or extreme. In the case of probabilistic models, this means that the main concern of the modeling work should be in the tails of probability distributions. Yet, when using parametric distributions, the tails are very sensitive to distribution assumptions and to distribution parameters. In the case of discrete distributions without assumption of the form of the distribution, the assessment of tails is still more difficult. These problems are common to all probabilistic approaches, and there have been innumerable attempts to overcome these problems by fuzzy set theory, rule-based systems, and many other approaches. However, the probabilistic approach (i.e., in risk analysis) is apparently becoming increasingly accepted in practice by administrative bodies and policy makers, and there is an apparent demand for efficient techniques for handling probabilistic information.

The belief network approach has, in different versions, been adopted in many fields (Bobrow 1993). It remains to be seen whether the same will occur in the natural resource and environmental sector. There are strong reasons for anticipating that there will be further studies using belief networks, which may include the possibility of performing two-directional, probabilistic computation on-line with apparently reasonable effort and accuracy, compatibility with Bayesian decision analysis and expected utility theory together with compatibility with deterministic management models, and of performing calculations from causes to effects and vice versa (Shachter & Heckerman 1987).

If a question were posed whether the proposed approach should replace state-of-the-art modeling approaches in cases such as the river water quality management example, our answer would be *not immediately*. In the study by Somlyódy et al. (1994), the used approaches include, for parameter estimation, Bayesian estimation and the Hornberger-Spear-Young approach (generalized sensitivity analysis). Both are well-tested and widely used approaches. For optimization of wastewater treatment plants, dynamic programming was used. It would be naive to propose that a novel, modestly investigated approach would immediately replace the existing methodology. For practical recommendations, robust, well-known approaches have many advantages. However, the importance of developing and testing novel, innovative approaches lies in the potential of finding completely new paths to solve problems within a discipline. The development of such paths is evidently a process likely to require more than just one case study. Even so, the proposed approach definitely has the potential for successful practical applications in near future.

Acknowledgments

This work was funded partly by IIASA and partly by the Ministry of Foreign Affairs of Finland. I am grateful to the supporting criticism of the colleagues at the IIASA Water Resources Project, especially to László Somlyódy, David Yates, Kenneth Strzepek, and Ilya Masliev. Many of the basic ideas have been influenced by Sakari Kuikka from Finnish Game and Fisheries Research Institute. I also want to thank my colleagues at the Helsinki University of Technology, and especially Petri Kylmälä who gave valuable comments on the concept of link strength parameter.

References

ADB 1990. *Environmental Risk Assessment – Dealing with Uncertainty in Environmental Impact Assessment.* ADB Environment Paper No. 7. Asian Development Bank, Manila Philippines.

Bobrow, D.G. 1993. Artificial Intelligence in perspective: a retrospective on fifty volumes of the *Artificial Intelligence Journal. Artificial Intelligence* 59, 5-20.

Caselton, W.F. & Luo, W. 1992. Decision making with inprecise probabilities: Dempster-Shafer theory with applications. *Water Resources Research* 28, 3071-3083.

De Jongh, P. 1988. Uncertainty in EIA. In: Wathern, P. (Ed.): *Environmental Impact Assessment: Theory and Practice*: 62-84. Routledge, London.

Gordon, J. & Shortliffe, E.H. 1985. A method of managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence* 26: 323-357.

Horwitz, E.J., Breese, J.S. & Henrion, M. 1988. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning* 2: 247-302.

Howard, R.A. 1968. The foundations of Decision Analysis. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, 211-219.

Howson, C. & Urbach, P. 1991. Bayesian reasoning in science. *Nature* 350, 371-374.

Keefer, D.N. & Bodily, S.E. 1983. Three-point approximations for continuous random variables. *Management Science* 29: 595-609.

Korn, G.A. & Korn, T.M. 1968. *Mathematical Handbook for Scientists and Engineers.* McGraw-Hill, New York.

Masliev, I. & Somlyódy, L. 1994. Uncertainty analysis and parameter estimation for a class of river dissolved oxygen models. *International Institute for Applied Systems Analysis, Laxenburg, Austria,* WP-94-9.

Miller, A.C. & Rice, T.R. 1983. Discrete approximations of probability distributions. *Management Science* 29: 352-362.

Morgan, M.G. & Henrion, M. 1990. *Uncertainty, A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, Cambridge, MA.

North, D.W. 1968. A tutorial introduction to decision theory. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, no. 3.

Pearl, J. 1986a. Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence* 29, 241–288.

Pearl, J. 1986b. On evidential reasoning in a hierarchy of hypotheses. *Artificial Intelligence* 28: 9-15.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan-Kaufmann, San Mateo, CA.

Raiffa, H. 1968. *Decision Analysis.* Addison-Wesley, Reading, MA.

Shachter, R.D. 1986. Evaluation of in_uence diagrams. *Operations Research* 34: 871-882.

Shachter, R.D. 1988. Probabilistic inference and influence diagrams. *Operations Research* 36: 589-604.

Shafer, G. 1990. Decision Making. In: Shafer, G. & Pearl, J. (Eds.): *Readings in Uncertain Reasoning:* 61-67. Morgan-Kaufmann, San Mateo, CA.

Shafer, G. & Pearl, J. (Eds.) 1990. *Readings in Uncertain Reasoning:* 61-67. Morgan-Kaufmann, San Mateo, CA.

Shachter, R.D. & Heckerman, D.E. 1987. Thinking backward for knowledge acquisition. *AI Magazine* 8, 55-61.

Shenoy, P.P. & Shafer, G. 1986. Propagating belief functions with local computations. *IEEE Expert*, Fall 1986: 43-52.

Somlyódy, L. 1993. Quo vadis water quality management in Central and Eastern Europe? *International Institute for Applied Systems Analysis, Laxenburg, Austria,* WP-93-68.

Somlyódy, L., Masliev, I., Petrovic, P. & Kularathna, M. 1994. Water quality management in the Nitra River Basin. *International Institute for Applied Systems Analysis, Laxenburg, Austria,* WP-94-40.

Szolovits, P. & Pauker, S.G. 1993. Categorial and probabilistic reasoning in medicine revisited. *Artificial Intelligence* 59, 167-180.

USEPA 1988. *Review of Ecological Risk Assessment Methods.* EPA–230–10–88–041. Washington D.C.

Varis, O. 1994. A belief network approach to modeling of environmental change: the methodology and prospects for application. *International Institute for Applied Systems Analysis, Laxenburg, Austria,* WP-94-40.

Varis, O., Kuikka, S. & Kettunen, J. 1993. Belief networks in fish stock assessment – the Baltic salmon case. *International Council for Exploration of the Sea,* C.M. D:13 (Mimeo).

Varis, O., Kuikka, S. & Taskinen, A. 1994. Modeling for water quality decisions: uncertainty and subjectivity in information, in objectives, and in model structure. *Ecological Modelling* 74, 91-101.

Wald, A. 1950. Statistical Decision Functions. Wiley, New York.

WCED 1987. *Our Common Future: Report of the World Commission on Environment and Development.* Oxford University Press, Oxford.

List of Symbols

The symbols used repeatedly in the text have been collected into this table.

e^*_i	evidence (normal distribution) to node i
$\mathbf{e}_i$	evidence vector (discrete distribution) to node i (one element: e_i)
$\mathbf{y}_i$	outcome vector of node i (one element: y_i)
$\mathbf{l}$	non-updated λ message vector (one element: l)
λ	updated λ message vector (one element: λ)
$\mathbf{p}$	non-updated π message (one element: p)
π	updated π message (one element: π)
$\mathbf{Bel}_i$	posterior distribution (belief)
$\mathbf{M}_{i\|j}$	link matrix from node i to node j
$\eta_{i\|j}$	link strength from node i to node j
$P(a)$	probability of a
$E(a)$	expected value of a
μ_a	mean of a
σ_a	standard deviation of a
cv_a	coefficient of variation of a
σ^2_a	variance of a
$\mathrm{var}(a)$	variance of a
$\mathrm{cov}(a, b)$	covariance of a and b
BOD	Biological oxygen demand (a water quality indicator)
DO	Dissolved oxygen (a water quality indicator)
NH4	Nitrogeneous BOD (a water quality indicator)

APPENDIX 2

Table A1. Hypothetical data on water quantity and water quality on the watershed.

			Reach									
		initial	1	2	3	4	5	6	7	8	9	10
Length (m)	*stream*		16500	8700	8100	10200	3900	3400	7100	25500	17800	41300
	tributary		10700	8200	1000	10100	1000	1000	6400	40000	20000	18400
Velocity (m/d)	*stream*	25000	26146	28199	37337	35758	23799	35913	29973	6108	7424	25909
	tributary		30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
Flow (m3/d)	*stream*	100000	127000	150000	175000	268700	270000	312100	345000	391000	404800	457000
	tributary		23000	25000	93700	1300	42100	32900	46000	13800	52200	50000
BOD (mg/l)	*stream*	13	11	13	11	15	11	18	17	8	9	5
	tributary		7	12	8	10	6	5	4	20	15	5
DO (mg/l)	*stream*	10	6	7	2.5	4	5	4	1.5	2.5	3	7
	tributary		10	9	6	10	7	9	8	5	7	4
NH4 (mg/l)	*stream*	4	3.7	2.5	8	3.5	4	7	6	5	3	2.2
	tributary		1	0.4	0.5	1	0.1	0.3	2	9	6	2

Table A2. Hypothetical data on investment costs due to different upgrading level alternatives of treatment plants (US$ 10^6).

Treatment					*Plant*					
alternative	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	1	0.1	0	1.5	0.5	1.2	2	0.5	0.1	0.5
3	1.3	0.15	3	2	1	9	9	1	0.2	3.9
4	7.8		2		2.5	12	15	7		
5			5		9	8.8	11			
6						9	14			
7						10.8	25			
8						14.4				

Table A3. Hypothetical data on operation and maintenance costs due to different upgrading level alternatives (US$ 10^6).

Treatment					*Plant*					
alternative	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0
1	0.35	0.07	1	0.7	0.4	0.6	1.1	0.3	0.17	0.17
2	0.46	0.1	1.2	0.8	0.52	0.7	1.3	0.4	0.22	0.22
3	0.55	0.12	1.5	0.95	0.58	0.6	1.2	0.5	0.25	0.34
4	0.7		1.4		0.7	0.7	1.4	0.6		
5			1.6		0.8	0.82	1.6			
6						0.9	1.9			
7						1.08	2.3			
8						1.45				

Table A4. Hypothetical data on BOD loads (kg/d) with different upgrading level alternatives.

Treatment					*Plant*					
alternative	1	2	3	4	5	6	7	8	9	10
0	1020	130	4250	2400	2808	3999	7824	1400	460	680
1	122.4	12	425	240	202.8	1419	3260	112	40	68
2	102	10	425	240	187.2	967.5	1956	112	30	68
3	81.6	10	375	240	156	322.5	782.4	84	24	34
4	40.8		375		140.4	322.5	586.8	56		
5			250		124.8	322.5	652			
6						258	489			
7						258	260.8			
8						129				

Table A5. Hypothetical data on NH4 loads (kg/d) with different upgrading level alternatives.

Treatment					*Plant*					
alternative		2	3	4	5	6	7	8	9	10
0	204	30	1000	480	390	619.2	1564.8	268.8	60	153
1	142.8	2	700	0	273	516	1304	134.4	4	81.6
2	27.2	0	100	0	273	516	1108.4	22.4	4	13.6
3	0	0	0	0	39	438.6	978	0	0	0
4	0		100		0	0	130.4	0		
5			0		0	438.6	228.2			
6						0	0			
7						0	0			
8						0				

Definition A3. For a belief network link $\mathbf{p}_{n|m} = \mathbf{M}_{n|m}\,\pi_m$ between two variables m and n, a link strength parameter $\eta_{n|m} \in [-1, 1]$ defines a $k \times k$ link matrix $\mathbf{M}_{n|m}$ so that the first, third, and fourth moments of $\mathbf{p}_{n|m}$ and π_m are equal. The second moment of $\mathbf{p}_{n|m}$ is obtained as $\sigma^2(p_{n|m}) = \eta_{n|m}^2 \sigma^2(\pi_m)$.

Theorem A3. The link strength parameter that fulfills the Definition A3 can be used to construct a link matrix $\mathbf{M}_{j|i}$ in the following way. For $\eta \geq 0$, the diagonal elements of $\mathbf{M}$ are obtained by

$$m_{q,r} = \frac{1}{k} + \eta\left(1 - \frac{1}{k}\right) \qquad , \qquad q = r = 1, \ldots, k \qquad \text{(A2-1)}$$

and the off-diagonal elements by

$$m_{q,r} = \frac{1}{k-1}\left[1 - \left[\frac{1}{k} + \eta\left(1 - \frac{1}{k}\right)\right]\right], \qquad q \neq r \qquad \text{(A2-2)}$$

For $\eta < 0$,

$$m_{q,r} = \frac{1}{k} + \eta\left(1 - \frac{1}{k}\right) \qquad , \qquad q = k - r \qquad \text{(A2-3)}$$

$$m_{q,r} = \frac{1}{k-1}\left[1 - \frac{1}{k} - \eta\left(1 - \frac{1}{k}\right)\right], \qquad q \neq k - r \qquad \text{(A2-4)}$$

Proof. For simplicity, let us denote $\mathbf{p}_{j|i} = \mathbf{M}_{j|i}\pi_j \triangleq \mathbf{p} = \mathbf{M}\pi$ and $\eta_{n|m} \triangleq \eta$. For $\eta \geq 0$:

$$\mathbf{p} = \mathbf{M}\,\pi$$

$$\Leftrightarrow \quad \mathbf{p} = \begin{bmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \\ \pi_k \end{bmatrix}$$

where $a = 1/k + (1 - 1/k)\eta$, and $b = (1 - a)/(k - 1)$

=> $$\mathbf{p}=\begin{bmatrix}\left[\frac{1}{k}+\eta\left(1-\frac{1}{k}\right)\right]\pi_1+\frac{1}{k-1}\left[1-\left[\frac{1}{k}+\eta\left(1-\frac{1}{k}\right)\right]\right](\pi_2+\pi_3+\ldots+\pi_n)\\ \left[\frac{1}{k}+\eta\left(1-\frac{1}{k}\right)\right]\pi_2+\frac{1}{k-1}\left[1-\left[\frac{1}{k}+\eta\left(1-\frac{1}{k}\right)\right]\right](\pi_1+\pi_3+\ldots+\pi_n)\\ \vdots\\ \left[\frac{1}{k}+\eta\left(1-\frac{1}{k}\right)\right]\pi_n+\frac{1}{k-1}\left[1-\left[\frac{1}{k}+\eta\left(1-\frac{1}{k}\right)\right]\right](\pi_1+\pi_2+\ldots+\pi_{n-1})\end{bmatrix}$$

=> $$p_i=\eta\left[\pi_i-\frac{1}{k}\sum_l \pi_l\right]+\frac{1}{k}\left[\pi_i-\sum_{j\neq i}\pi_j\right],\quad i=1,\ldots,k;\ j=1,\ldots,k,\ l=1,\ldots,k$$

1° First moment $E(p)$

$$E(p)=\frac{1}{k}\sum_i p_i$$

=> $$E(p)=\frac{1}{k}\sum_i\left[\eta\left[\pi_i-\frac{1}{k}\sum_l \pi_l\right]+\frac{1}{k}\left[\pi_i-\sum_{j\neq i}\pi_j\right]\right]$$

=> $$E(p)=\frac{1}{k}\left\{\eta\left[\left(1-\frac{1}{k}\right)\pi_1-\frac{1}{k}(\pi_2+\pi_3+\ldots+\pi_n)\right]+\frac{1}{k}(\pi_1-\pi_2-\pi_3-\ldots-\pi_n)\right\}$$
$$+\frac{1}{k}\left\{\eta\left[\left(1-\frac{1}{k}\right)\pi_2-\frac{1}{k}(\pi_1+\pi_3+\ldots+\pi_n)\right]+\frac{1}{k}(\pi_2-\pi_1-\pi_3-\ldots-\pi_n)\right\}$$
$$\ldots$$
$$+\frac{1}{k}\left\{\eta\left[\left(1-\frac{1}{k}\right)\pi_n-\frac{1}{k}(\pi_1+\pi_2+\ldots+\pi_{n-1})\right]+\frac{1}{k}(\pi_n-\pi_1-\pi_2-\ldots-\pi_{n-1})\right\}$$

=> $$E(p)=\frac{1}{k}\left\{\eta\left(\sum_i\pi_i-\sum_i\pi_i\right)\right\}+\frac{1}{k}\left(\sum_i\pi_i+(k-1)\sum_i\pi_i\right)=\frac{1}{k}\sum_i\pi_i=E(\pi)$$

Q.E.D.

2° Second moment $\sigma^2(p) = E(p - E(p))^2$

$$\sigma^2(p)=\frac{1}{k-1}\sum_i(p_i-E(p))^2$$

=> $$\sigma^2(p)=\eta^2\sigma^2(\pi)$$

The proof is analogical to that in 1°

3° Third moment $\sigma^3(p)$

$$\sigma^3(p) = \frac{1}{k-1}\sum_i (p_i - E(p))^3$$

=> $\sigma^3(p) = \sigma^3(\pi)$

The proof is analogical to that in 1°

4° Fourth moment $\sigma^4(p)$

$$\sigma^4(p) = \frac{1}{k-1}\sum_i (p_i - E(p))^4$$

=> $\sigma^4(p) = \sigma^4(\pi)$

The proof is analogical to that in 1°

For $\eta < 0$,

$$\mathbf{p} = \mathbf{M}\,\pi$$

<=> $$\mathbf{p} = \begin{bmatrix} b & \cdots & b & a \\ b & \cdots & a & b \\ \vdots & \iddots & \vdots & \vdots \\ a & \cdots & b & b \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \\ \pi_k \end{bmatrix}$$

where $a = 1/k + (1 - 1/k)\eta$, and $b = (1 - a)/(k - 1)$

The proofs are analogical to $\eta \geq 0$.