Parameterising Bayesian Networks: A Case Study in Ecological Risk Assessment.

Owen Woodberry, Ann Nicholson, and Kevin Korb
School of Computer Science and Software Engineering
Monash University, Clayton, Victoria 3800, Australia
{owenw,annn,korb}@csse.monash.edu.au

Carmel Pollino
Water Studies Centre
Monash University, Clayton, Victoria 3800, Australia
carmel.pollino@sci.monash.edu.au

Abstract

Most documented Bayesian network (BN) applications have been built through knowledge elicitation from domain experts (DEs). The difficulties involved have led to growing interest in machine learning of BNs from data. There is a further need for combining what can be learned from the data with what can be elicited from DEs. In this paper, we propose a detailed methodology for this combination, specifically for the parameters of a BN. We illustrate the techniques using a case study of an ecological risk assessment (ERA) problem, specifically the Goulburn Catchment (Victoria, Australia) ERA project.

1 Introduction

Bayesian networks (BNs) are graphical models for probabilistic reasoning, which are now widely accepted in the AI community as intuitively appealing and practical representations for reasoning under uncertainty. A BN is a representation of a joint probability distribution over a set of statistical variables. It has both a qualitative aspect, the graph structure, and a quantitative aspect, marginal and conditional probabilities. The structure is a directed acyclic graph (DAG) and formally represents the structural assumptions of the domain, i.e., the variables comprising the domain and their direct probabilistic dependencies, which are typically given a causal interpretation.

The quantitative aspect associates with each node a conditional probability table (CPT), which describes the probability of each value of the child node, conditioned on every possible combination of values of its parents. Given both the qualitative and the quantitative parts, probabilities of any query variables posterior to any evidence can be calculated [22].

Most reported BN applications to date (including medical and other diagnosis, planning, monitoring and information retrieval - see [16, Ch.5] for a recent survey) have been built through knowledge elicitation from domain experts (DEs). In general, this is difficult and time consuming [9], with problems involving incomplete knowledge of the domain, common human difficulties in specifying and combining probabilities, and DEs being unable to identify the causal direction of influences between variables. Hence, there has been increasing interest in automated methods for constructing BNs from data (e.g., [27, 12]).

Thus far, a methodology and associated support tools for Knowledge Engineering Bayesian Networks (KEBN) are not well developed. Spiral, prototype-based approaches to KEBN have been proposed (e.g., [17, 16]), based on successful software development processes [4, 2]. However, these provide little guidance on how to integrate the knowledge engineering of the qualitative and quantitative components or again on how to combine knowledge elicitation from DEs and automated knowledge discovery methods. While there have been attempts at the latter, they remain rudimentary (e.g., [21, 19]).

Here we present a more detailed methodology, based on the spiral prototype model, for knowledge engineering the quantitative component of a BN. Our methodology explicitly integrates KE processes using both DEs and machine learning, in both the parameter estimation and the evaluation phases. The methodology was developed during the knowledge engineering of an ecological risk assessment application (Section 2), hence we use illustrative examples from this case study throughout.

2 Ecological Risk Assessment Application

The case study for this paper is knowledge engineering for Ecological Risk Assessment (ERA). In the ERA domain, there is an increasing need for environmental decision support tools that are able to model complex ecosystems in an integrated framework, acknowledging that uncertainties in input in-formation and predictive outputs exist. Currently, few tools meet these requirements. Those that are available are often highly complex, poorly tractable and not particularly user friendly.

The objective of the Goulburn Catchment (Victoria, Australia) ERA project is to develop a model to determine the effects of alternative management actions and associated environmental conditions on the native fish community in the Goulburn Broken Catchment, Victoria, Australia. There is evidence that native fish communities have declined substantially since barriers (dams and weirs) were placed in the catchment [5, 10, 11]. Although the major factors for the decline have been identified previously [10], changes to fish communities have not been measured consistently over time.

BNs were found to meet the domain's needs, being: explicit representation of the causal interactions in the model; representation of uncertainty; ability to combine domain information from DEs and the monitoring data; ability to incorporate information from multiple scales; and provision of model output that is directly applicable to risk management.

2.1 Background: Goulburn Catchment

The main stem of the Goulburn Catchment, the Goulburn River, is the largest tributary of the Murray-Darling Basin in the State of Victoria (Australia). The lowland Goulburn River extends from Eildon to its confluence with the Murray River at Echuca. Many rivers and creeks enter the 436 km lowland stretch of the Goulburn River. The headwaters of the Goulburn River flow into Lake Eildon. Water released from Lake Eildon is delivered 218 km downstream to Goulburn Weir. From Goulburn Weir, outflows are to the lower Goulburn River and three irrigation channels. There is evidence that native fish communities in the Goulburn Catchment have declined over the past 100 years (see [23]). Four major factors have been identified as influencing native fish abundance and diversity in the Goulburn River, being water quality, flow alterations, instream habitat and biological interactions. Although the processes and interactions between these factors and their link to native fish decline are broadly understood, quantitative models to assist in environmental management do not exist.

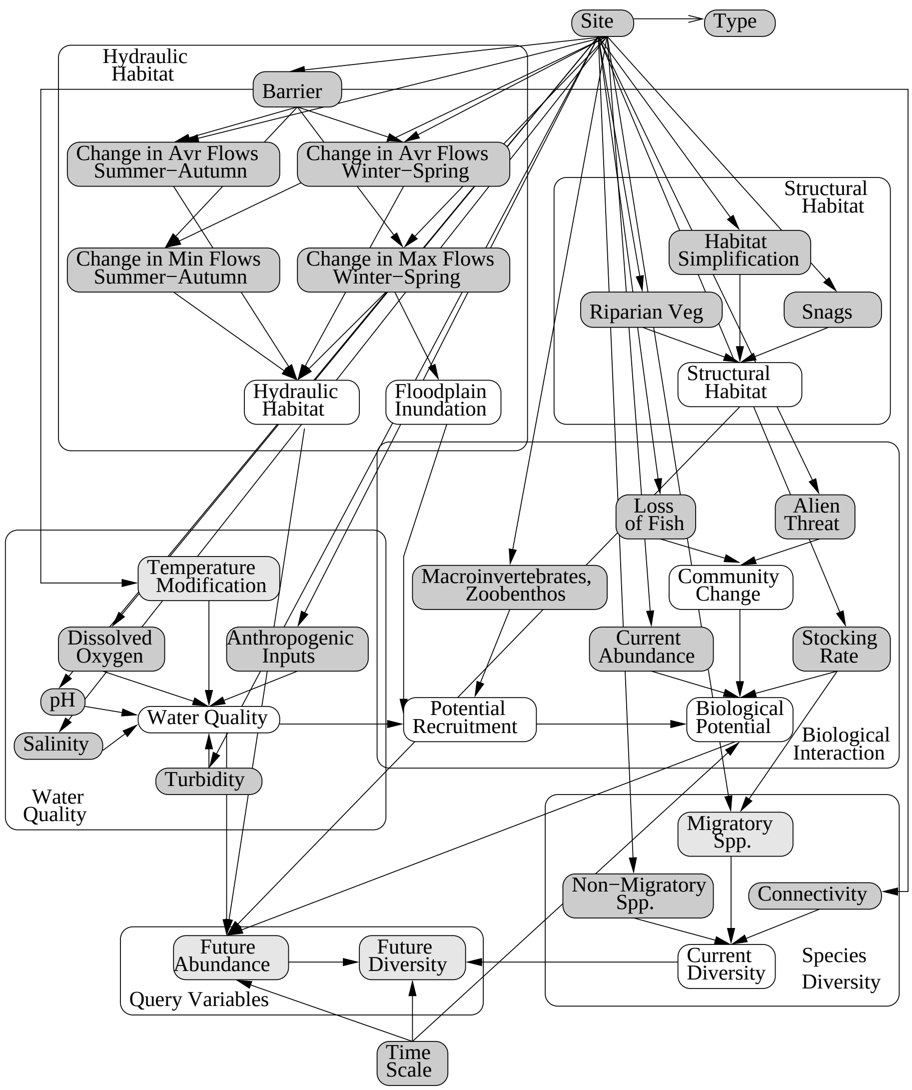

Figure 1: Prototype ERA BN

2.2 Graphical Structure

One version of the BN for this application developed by the main domain expert (DE), is shown in Figure 1.

The BN structure is based on a comprehensive conceptual model developed by DEs. This ERA model consists of five interacting components – water quality, hydraulic habitat, structural habitat, biological potential and species diversity – and has two query variables: Future Abundance and Future Diversity, of the native fish communities For the purposes of this paper, variable names have been simplified. This model was built prior to interaction with the BN technology experts who subsequently acted as knowledge engineers for the project.

The query variables were also established in collaboration with DEs and are based on a very pragmatic management need in the Goulburn Catchment, being to assess what conditions are required to establish sustainable native fish communities. Given that no native fish recruitment data was available, the surrogate endpoints used were based on what information was available.

To assess the impacts of human-related activities on native fish communities, it was important to begin constructing the BN by establishing linkages between processes and activities of importance. Clearly, the Goulburn Catchment is a highly complex system with multiple factors interacting and influencing fish communities. The empirical relationships within and between chemical, physical and biological system components have not been previously characterized.

Table 1: ERA Data: Number of cases for each site, and any additional missing variables. Sites are grouped into geographical regions.

Region	Site	No.Cases	Missing data %	Fully Missing Variables
Upper Main	G_Eild	97	20%	
	G_Alex	20	7.5%	
	G_Yea	5	26.9%	Temp, Turb, MinSummer
	G_Trawool	4	26.9%	MaxWinter, Food,FutureAbundance, Future Diversity
Upper Tributary	Rubi	27	11%	
	Tagg	25	13.7%	Fishing, Stocking
	Ach	90	32.5%	Food, Fishing
	Murrin	83	26.4%	Fishing
	Yea	24	11.4%	Fishing
	King	15	11%	Fishing
	Sunday	8	16.8%	Fishing, FutureAbundance,FutureDiversity
	Hughes	10	18.5%	Salinity, Turb, Fishing
Mid Main	G_LkNag	12	11.5%	Food, FutureAbundance,FutureDiversity
Lower Main	G_Murch	6	7.7%	FutureAbundance, FutureDiversity
	G_Shep	14	10.4%	Food
	G_McCoy	272	26%	FutureAbundance, FutureDiversity
	G_Undera	4	11.5%	Turb, FutureAbundance,FutureDiversity
	G_Echuca	3	19.2%	Salinity, Turb, Food, FutureAbundance, FutureDiversity
Lower Tributary	Pranjip	8	15.4%	Turb, Food, FutureAbundance,FutureDiversity
	Crei_Bran	12	18.9%	Salinity, Turb, Fishing
	Castle	6	11.5%	Fishing, FutureAbundance,FutureDiversity
	Sevens	31	14.8%	
Broken	Broken	173	19.4%	
Total		949		

We received data from 23 sites which are further aggregated into 6 regions, ranging from 3 to 272 cases, with 949 cases in total (see Table 1). For the purpose of parameter learning each case was counted twice to match cases with one- and five-year projections of future abundance.

In Figure 1, the variables for which there was data are indicated by shading, although some sites have data missing for particular variables (see Table 1, last column). The number of cases for each site, plus the percentage of data missing, is also shown in Table 1, in columns 3 and 4 respectively. Where variables had only limited data available, parameters were initially elicited from DEs (see Section 4). These variables are indicated by lighter shading.

The development of the model structure is not described in any more detail in this paper, other than to acknowledge that it was undertaken via an iterative process.

3 Quantitative Knowledge Engineering Methodology

A possible methodology for the quantitative knowledge engineering of BNs is outlined in Figure 2. This method illustrates possible flows (indicated by arrows) through the different KE processes (rectangular boxes), which will be executed either by humans (the DE and the knowledge engineer, represented by clear boxes) or computer programs (shaded boxes). Major choice points are indicated by hexagons.

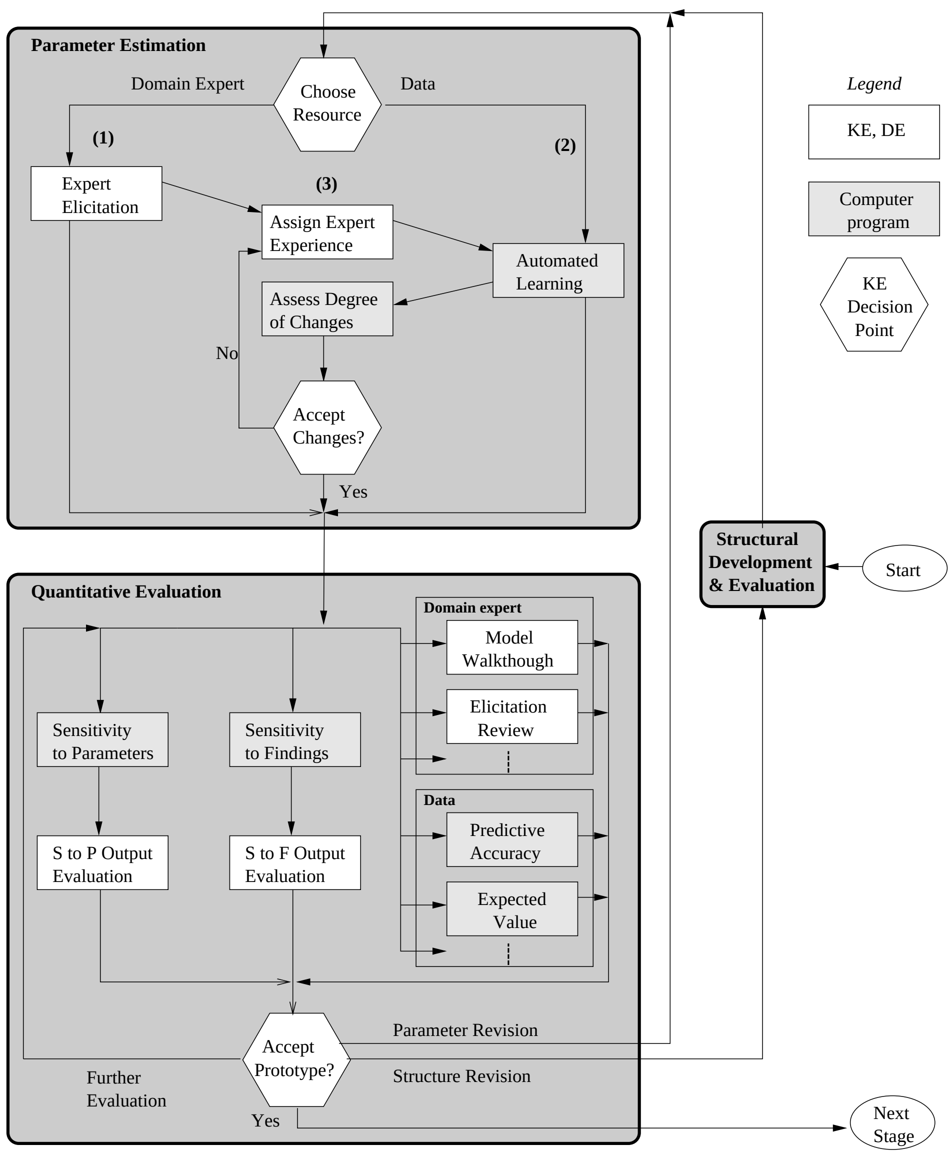

Figure 2: Quantitative Knowledge Engineering Methodology (see text for further explanation)

The initial stage in the development spiral is **Structural Development and Evaluation**, which on the first iteration will produce an unparameterized causal network; a network structure must exist prior to parameterization and may need to be reconsidered after evaluation We do not describe this process in any detail, however it should also proceed in an iterative fashion, using techniques such as those in [3] We do note that the d-separation properties of the qualitative network — a graph-theoretic definition of which nodes "cut off" some nodes from others — need to reflect the actual probabilistic independencies in the data (see [22]). The modelling choices made at this qualitative stage can have a large impact on the number of parameters that must be estimated, since the size of the CPT is a function of joint state space across all parents. Subsequently, for earlier prototypes it is preferable to simplify the network by focusing on the most important relationships and reducing the state spaces of parent variables by making coarser distinctions between categories.

Once a BN structure has been established, the next step is **parameter estimation**, involving specifying the CPTs for each node. Figure 2 shows that the parameter estimates can be elicited from DEs (1),[1] or learned from data (2) or, as proposed here, generated from a combination of both sources (an example is shown in path 3). In early prototypes the parameter estimates need not be exact, and uniform distributions can be used if neither domain knowledge nor data are readily available. A detailed description of the parameter estimation process is provided in Section 4 below.

The second major aspect of quantitative knowledge engineering is **quantitative evaluation**. Evaluative feedback can be generated using either DEs or data or both, as we have done here. When data is available, several measures can be used to evaluate BNs, including predictive accuracy, expected value computations and information reward. DE evaluation techniques include elicitation reviews and model walkthroughs (see Figure 2). Another kind of evaluation is *sensitivity analysis.* This involves analysing how sensitive the network is, in terms of changes in updated probabilities of some query nodes to changes in parameters and inputs. Measures for these can be computed automatically using BN tools (shown as **Sensitivity to Parameters** and **Sensitivity to Findings** processes, in Figure 2), but these need to be evaluated by the DE in conjunction with the KE. A detailed description of sensitivity analysis is given in Section 5.

3.1 ERA Case Study

The domain expert conducted two stakeholder workshops with management and fishery experts. The first workshop concentrated on the BN casual structure, and the second on the parameter estimates. The domain expert combined this source of information with information from the literature to specify a preliminary causal structure and parameters for all child variables. The remaining root nodes were given a uniform distribution. This provided a basic initial BN, which was then developed further in conjunction with the knowledge engineers, refining the parameters using data and the evaluation techniques described above.

Because the ERA project had the involvement of a domain expert developer as well as input from other domain experts it was appropriate to, initially, specify elicited parameter values. Although, due to the lack of knowledge, and confidence, of the domain experts it was also relevant to incorporate the limited observational data available. As both sources were limited and inadequate by themselves, a special focus was given to the combination of sources to develop a useful and informative model.

[1] This can also include the domain literature as a source of parameter estimates.

4 Parameter Estimation

4.1 Elicitation from experts

During **expert elicitation** the DEs provide or refine estimates of the BN parameters. Direct elicitation employs such questions as *"What is the probability that variable **A** takes this state given these parent values?"* Alternatives are to use frequencies, odds, or qualitative elicitation, using terms such as 'high' or 'unlikely', with the mapping to actual probabilities calibrated separately. The VE (Verbal Elicitor) tool [14] was designed to support such qualitative elicitation.

In addition to eliciting precise parameters, it can also be useful to elicit an acceptable range for the parameter. As many are familiar with 95% confidence intervals from statistics, DEs might be comfortable reporting intervals having a 95% chance of capturing the desired parameter, although other ways of specifying a range of values are equally legitimate. Such intervals can be used during later evaluation to identify parameters needing further attention, as we shall see.

4.1.1 ERA Case Study

Variables were split into two groups: those with data were initially given uniform probability distributions; the remainder (see Figure 1) were first fully elicited from DEs. DEs were asked to report their confidence in these estimates, which was categorized as either low or high. We note that DEs tended to be more confident estimating variables pertaining to the physical and chemical relationships in the system and less so with the biological relationships. In this study, the elicited confidence applied to the node, i.e., to the whole CPT, rather than to individual parameters; this need not be the case in general.

4.2 Learning parameters from data

When data is of good quality and voluminous, estimating parameters from the data is clearly preferable. Many techniques are available for this. The simplest is the Lauritzen & Spiegelhalter method (L&S) [18], using frequency counts of child states given each possible parent instantiation, although problems arise when the data lacks coverage of many states. Algorithms such as expectation maximization (EM) [8] and gradient descent (GD) [20] are built into many BN tools to deal with missing data by finding parameterizations which yield the greatest likelihoods given the data available.

Problems with incomplete data can be ameliorated also by incorporating other sources of information for parameters, such as expert knowledge, before automated learning. The combination of elicitation and data-based parameterization requires the elicited information to be weighted relative to the data available. In Figure 2 this is done in the **Assign Expert Experience** process, where an experience weighting is assigned to the expert parameter estimates, based on the confidence in the estimates obtained during expert elicitation. These are then treated as equivalent to the size of a hypothetical initial data sample (the equivalent sample size, or ESS).

4.3 Comparing expert elicited and automated parameterizations

After incorporating the data in parameter estimation, the next step is to compare the new with the original parameterization. In Figure 2 we consider this to be an automated process, **Assess Degree of Changes**. As mentioned above, during parameter elicitation an acceptable range of values can also be elicited. Any parameters estimated from the data to be outside this range should be flagged for attention. In this study, however, we used an alternative method for comparing the parameterizations, namely the Bhattacharyya distance [1] between the two probability distributions. This distance is computed for each possible combination of parent values; higher distances between conditional distributions trigger further attention. The DE must then assess whether these flagged parameter refinements obtained after automated learning are acceptable (in the **Accept Changes** decision point in Figure 2). If not, we undertook an iterative investigation of different mappings of the expert experience into equivalent sample sizes.

4.3.1 ERA Case Study

Of the L&S, EM and GD automated learning methods available in the Netica BN software [20], the EM method was selected, since the L&S method was not very useful with many parent instantiations missing in the data and the GD method was susceptible to local maxima. Automated learning trials were then carried out using EM in order to investigate the effects of different weightings of expert elicited CPTs. A pre-trial with the L&S method was used for comparative purposes.

Table 2: Nodes whose CPTs were first expert elicited, with the different experience weightings used for trials of the EM automated learning method.

	H=10,M=5			H=20,M=10			Combined	
	Trial No.			Trial No.			Trial No.	
Node	1	2	3	1	2	3	4	5
Water Quality (P)	10	15	18	20	25	25	25	24
Hydraulic Habitat (P)	10	15	18	20	25	25	25	24
Structural Habitat (P)	10	7	4	20	15	10	1	1
Biological Potential (B)	5	2	4	10	5	5	5	5
Temperature Modification (P)	10	5	1	20	15	10	1	1
Community Change (B)	5	1	1	10	5	1	1	1
Floodplain Inundation (P)	10	5	1	20	15	10	1	1
Potential Recruitment (B)	5	1	3	10	5	2	3	3
Connectivity (P)	10	10	10	20	17	14	12	12
Migratory spp (B)	5	10	15	10	15	15	15	16
Current Diversity (B)	5	1	1	10	7	4	1	2
Future Abundance (B)	5	2	4	10	7	4	5	6
Future Diversity (B)	5	5	5	10	5	5	5	5
Remaining Nodes	0	0	0	0	0	0	0	0

Note: P indicates variable part of physical process, B part of biological.

A series of trials were conducted (see Table 2) using the EM method to refine the parameters of all nodes. Each trial used a series of experience weightings. Each of these trial EM parameterizations was compared, using the Bhattacharyya distance, with the L&S BN, and an assessment was made as to whether the degree of change was acceptable. If the change was deemed unacceptably large, the ESS was increased, while if there was no or minor changes, the ESS was decreased. This assessment process was iterated, comparing the new EM parameterization with the L&S parameterization and setting a new ESS value, W_{i+1}, using Algorithm 1.

5 Quantitative Evaluation

After parameterization of the BN, the second major aspect of quantitative knowledge engineering is evaluation, which guides further iterations of BN development.

Algorithm 1 Adjusting ESS (with ESS changes from ERA case study)

Loop until **ESS** values converge
Parameterize network with current **ESS** values
  Switch
    Case changes unrealistic: $W_{i+1} \leftarrow W_i$+`upLarge (5)`
    Case would allow greater changes : $W_{i+1} \leftarrow W_i$+`upSmall (3)`
    Case little OR no change: $W_{i+1} \leftarrow W_i$+`downLarge (5)`
    Case insignificant change: $W_{i+1} \leftarrow W_i$-`downSmall (3)`
    Case changes become unrealistic: $W_{i+1} \leftarrow W_{i-1}$+`bounceup (2)`
    Case changes disappear: $W_{i+1} \leftarrow W_{i-1}$-`bouncedown (2)`
    Case final trials, small adjustments required: $W_{i+1} \leftarrow W_{i-1}\pm$`tweak (1)`
    Case changes acceptable: $W_i \leftarrow W_i$
End Loop

5.1 Data-Driven Evaluation

When data is available, it can be used for evaluation. Where the data is also being used to learn the structure or the CPTs, it is necessary to divide it into training data and test data, so that evaluation is not done with the very same data used for learning. The most common method of evaluation is to determine the *predictive accuracy* of the BN, which measures the frequency with which the modal node state (that with the highest probability) is observed to be the actual value.

5.1.1 ERA Case Study

As the ERA data was limited, automated evaluation was of little use. Predictive accuracy was used to test the situations that were available in the data The error rate of the query nodes, Future Abundance and Future Diversity, were only 5.8% and 0%, respectively. This was not very informative as there were only 10 High Future Abundance cases and 0 High Future Diversity cases. Thus, they were equal to the percentage of High case for each query node.

Other, less formal, data evaluation was conducted as follows. Current abundance and diversity at each of the sites under existing environmental conditions was plotted against model output probabilities of Future Abundance (see Figure 3) and Future Diversity (see Figure 4). The trends between real data and BN predictions are well maintained throughout the catchment. Output from both the L&S and EM learned models are included for comparative purposes, showing a general improvement in the predictive output of the EM model and hence our automated learning methodology.

5.2 Evaluation by domain expert

Even when adequate data is available, it is important to involve the DE in evaluation. If expert elicitation has been performed, a structured review of the probability elicitation is important.[2] This procedure could involve: comparing elicited values with available statistics; comparing values across different DEs and seeking explanation for discrepancies; double-checking cases where probabilities are extreme (i.e., at or close to 0 or 1), or where the DEs have indicated a low confidence in the probabilities when originally elicited.

ERA Case Study: The domain expert developer conducted a semi-formal model walkthrough with management and ecology experts, with positive feedback. It was recognized by each that more

[2]Note that the review of the variable and value definitions, and review of the structure, through both the implications of the d-separation dependencies and independencies and prior knowledge about time and causality, would occur during structural development, rather than this quantitative evaluation phase.

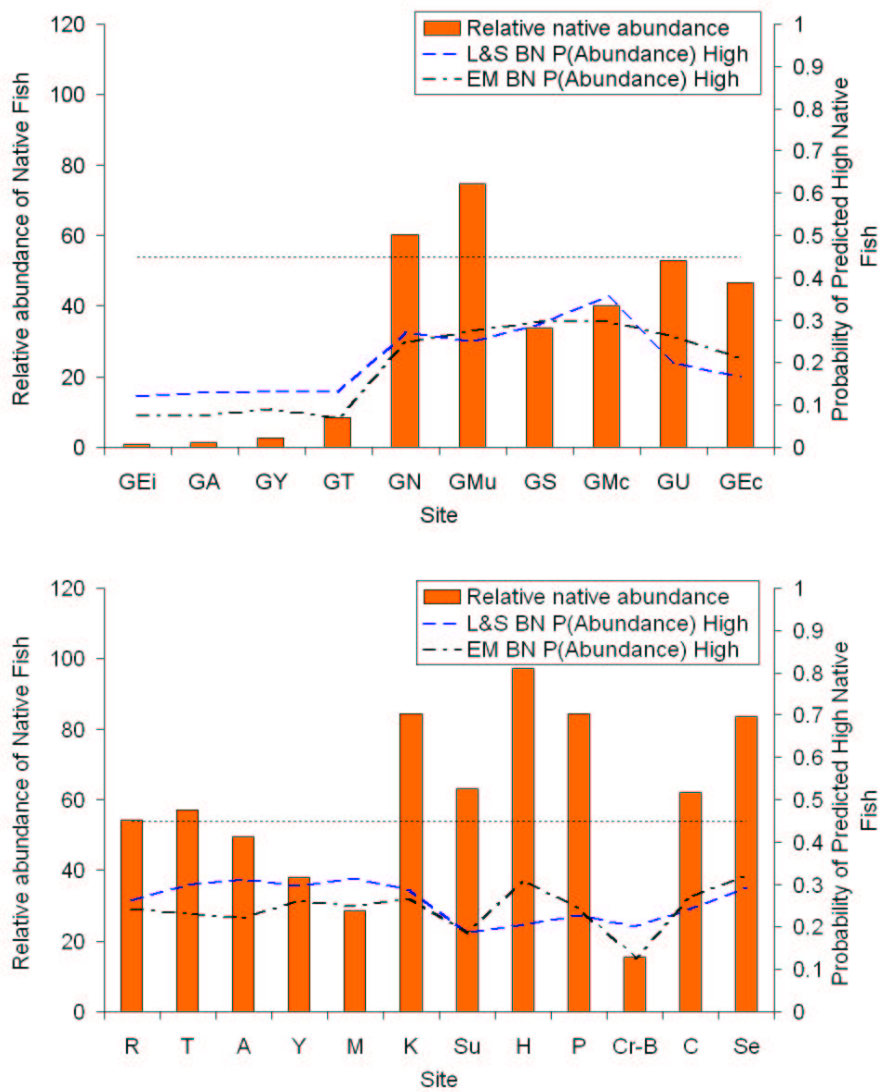

Figure 3: Relative Abundances of Native Fish at (a) Goulburn river sites (b) Tributary sites, (> 1970 fisheries data) vs. BN Predicted Abundance of 'High' (> 54) Native Fish at the same sites

case data was needed, especially with healthy fish populations, to strengthen the model. The lack of case data was recognized at the commencement of the study and the BN approach was used so that expert opinion could compensate.

We now review two different types of sensitivity analysis, discuss how we adapted them into algorithms suitable for our purposes, and describe examples from the case study. One type of sensitivity study looks at how the BN's posterior distribution changes under different observed conditions, in a "sensitivity to findings" study. The other looks at how the model's distribution changes when particular parameters are altered. Curiously, researchers thus far appear to have employed one or the other of these, but not both in any one study (e.g., [7, 17, 25]). Both are needed for a careful investigation of the properties of a network.

5.3 Sensitivity to Findings analysis

The properties of d-separation can be used to determine whether evidence about one variable may influence belief in a query variable. It is possible to measure this influence and rank evidence nodes

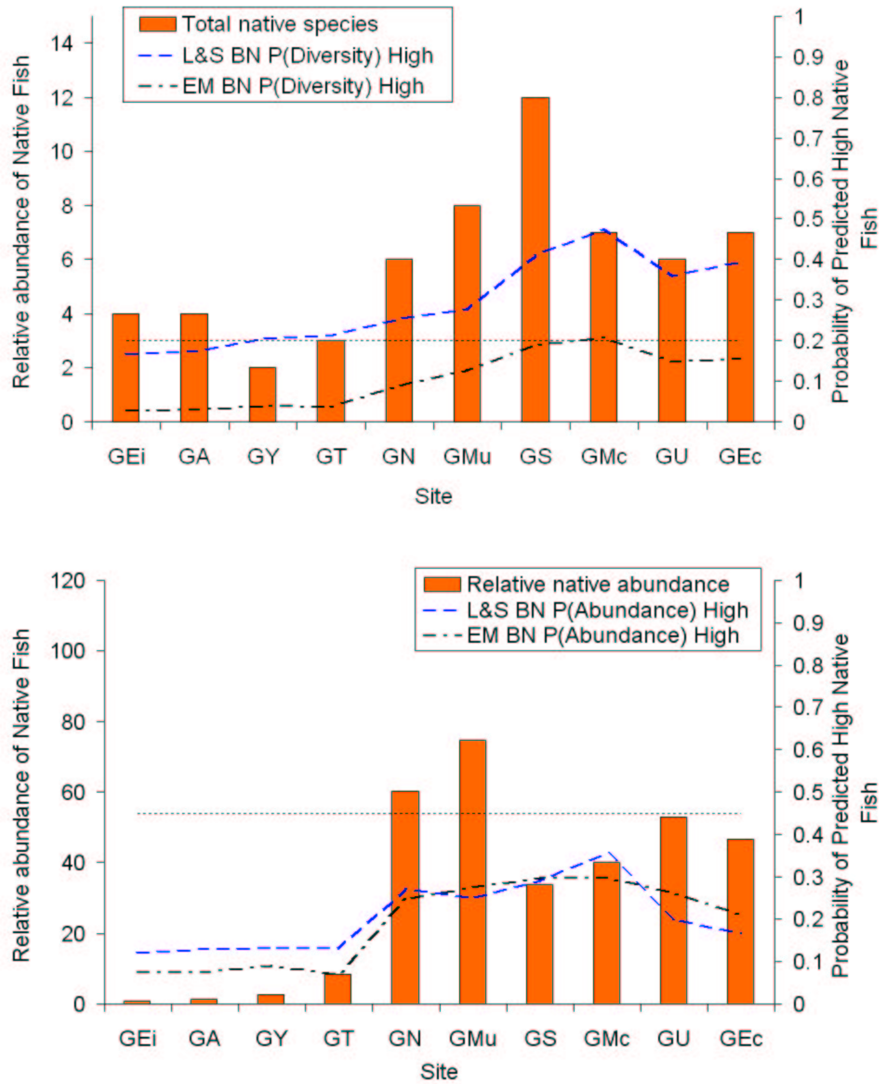

Figure 4: Total Number of Species of Native Fish at (a) Goulburn river sites (b) Tributary sites, ($>$ 1970 fisheries data) vs. Predicted Diversity of 'High' Native Fish (> 3) at the same sites

by how much of an effect they have. This information can be used to provide guidance for collecting the most informative evidence or as a check on whether the model reflects the DE's intuitions.

Sensitivity to findings can be quantified using two types of measures, entropy and mutual information. **Entropy**, $H(X)$, is commonly used to evaluate the uncertainty, or randomness, of a probability distribution $H(X) = -\sum_{x \in X} P(x) \log P(x)$. Measuring the effect of one variable on another is referred to as the **mutual information (MI)** $I(X|Y) = H(X) - H(X|Y)$,

We have implemented this type of sensitivity to findings (see Algorithm 2). Our algorithm computes and displays both the entropy of a specified query node and the ranked mutual information values for a specified set of interest nodes, given a set of evidence for some other observed nodes. The user can subsequently investigate how changes to the evidence will affect the entropy and MI measures. This process allows the DE to identify whether a variable is either too sensitive or insensitive to other variables in particular contexts, which in turn may help identify errors in either the network structure or the CPTs.

Algorithm 2 Sensitivity to Findings

```
loop
  Compute entropy of query node and mutual information values of interest nodes
  Display entropy of query node
  Rank and Display mutual information values
  Prompt user for action
  if Action = Set Node Finding then
    Prompt user for node name and value
    Enter node finding
  else if Action = Back up then
    Remove last node finding
  else if Action = Save Report then
    Save sensitivity analysis results to file
    break Loop
  end if
end loop
```

5.3.1 ERA Case Study

Application of the sensitivity to findings algorithm showed that the query variable, Future Abundance, in the absence of other evidence, is most sensitive to Future Diversity, followed by Water Quality (WQ) (see Table 3). When findings for WQ were entered into the network, the sensitivity measures changed, and the ranking of variables changed. When WQ = Low, evidence about other variables have less impact on the Future Abundance node, demonstrating the dominating effect of a low water quality in the system. Alternatively, when WQ = High, some of the remaining variables became more influential. These observations agreed with the DE's understanding of the system.

5.4 Sensitivity to Parameters analysis

For some time, the belief was common in the BN community that BNs are generally insensitive to inaccuracies in their parameters. Hence, the standard complaint that BNs required the specification of precise probabilities ("where will the numbers come from?") was met with the rejoinder that "the numbers don't really matter." This belief was based on a series of experiments described in [24], which was further elaborated on in [13]. Clearly, however, the inference that insensitivity is general is wrong, since it is easy to construct networks that are highly sensitive to parameter values. Indeed, more recent research has located some practical networks that are highly sensitive to imprecision in particular parameters [15, 26, 6].

Identifying sensitive parameters in a BN is important for focusing the knowledge engineering effort, for it will focus effort in refining parameterization on those values which have the biggest impact on the target variables. How best to identify these sensitivities remains a current research topic.

Sensitivity analysis could be performed using an empirical approach, by altering each of the parameters of the query node and observing the related changes in the posterior probabilities of the target node. However, such a straightforward analysis can be extremely time consuming, especially on large networks. Coupé and Van der Gaag [7] address this difficulty by first identifying a "sensitivity set" of variables given some evidence. These are those variables which can potentially change, meaning the remaining variables can be eliminated from further analysis. The sensitivity set can be found using an adapted d-separation algorithm. Coupé and Van der Gaag also demonstrated that the posterior probability of a state given evidence under systematic changes to a parameter value can be given a functional representation, either linear or hyperbolic.

Table 3: Sensitivity to Findings Analysis performed on Future Abundance

	No Evidence		Water Quality = Low		Water Quality = High	
Entropy of Future Abundance	0.762080		0.379315		0.878575	
Node	MI	Rank	MI	Rank	MI	Rank
Future Diversity	0.086744	1	0.039833	1	0.097424	1
Water Quality	0.055666	2	-	-	-	-
Hydraulic Habitat	0.031075	3	0.005350	2	0.029836	3
Biological Potential	0.030830	4	0.000189	21	0.056507	2
Site	0.027754	5	0.002104	3	0.006124	4
Barrier	0.022200	6	0.001866	6	0.002228	11
Type	0.022190	7	0.001890	5	0.003686	6
Temperature	0.021858	8	0.001827	7	0.000999	18
Avr Summer	0.019413	9	0.001919	4	0.003406	7
Min Summer	0.011461	10	0.001400	9	0.003406	8
Avr Winter	0.009974	11	0.001524	8	0.002454	9
Max Winter	0.008584	14	0.001235	10	0.002357	10
Current Abundance	0.004246	16	0.000148	22	0.004836	5

Algorithm 3 Sensitivity to Parameters

```
loop
  Prompt user for action
  Switch
        Case Set Interest Node
           Prompt user for node name
        Case Set Test Node  Parameter
           Prompt user for node name
           Prompt user for parent set (CPT row)
           Prompt user for test node state (CPT column)
        Case Set Evidence Set
           Prompt user for evidence set
           Set node findings
        Case Display Plot
          if test node in sensitivity set then
              Compute sensitivity function coefficients
              Display sensitivity function to screen
          end if
        Case Find Sensitive Parameters
           if test node in sensitivity set then
              for each parameter in test node do
                 Compute sensitivity function coefficients
                 Compute sensitivity value at parameter value
                 Display sensitivity And parameter values
              end for
           end if
end loop
```

We have implemented this type of sensitivity to parameters (see Algorithm 3). When a particular evidence instantiation is set, our algorithm identifies the type of sensitivity function for the parameters by checking whether the query node has any observed descendant nodes. Once the sensitivity function is determined for a parameter, its coefficients can be computed (see Algorithm 4). If the plotted sensitivity function does not behave as the DE expects (its slope, direction or range is unexpected), then this could indicate errors in the network structure or CPTs.

The revised normalized probability distribution of the test node is set by first selecting a new value, P_{new} for the parameter under investigation, P_j. The remaining parameters, P_i, are normalized to retain relative values by the updating function,

$$P_i \Leftarrow P_i \times \frac{1 - P_{new}}{1 - P_j}, i \neq j \tag{1}$$

before the parameter under study is updated,

$$P_j \Leftarrow P_{new} \tag{2}$$

Algorithm 4 Algorithm to find Sensitivity Functions

```
Loop 2 times for linear, 3 times for hyperbolic
    Set new normalized probability distribution in test node
    Compile network
    Get node belief in interest node
    Store parameter and belief pair
    Restore old probability distribution
End Loop
Solve for coefficients of sensitivity function
```

5.4.1 ERA Case Study

The most sensitive parameter identified in the ERA model was $P(FutureAbundance = Low|Time\_Scale = One\_year, WQ = Low, StrucHab = Low, BiolPoten = Low, OverallFlow = ExtIncrease)$ for the posterior probability $P(FutureAbundance|Site = G\_Eild, Time\_Scale = One\_year)$ (see Figure 5). This observation agreed with the DE evaluation of the system, in the magnitude, direction and range of the visual representation of the function.

5.5 KE Decision: Accept prototype

Quantitative evaluation can be used to identify problems with the BN structure and parameters. After the model has been evaluated using a particular technique, the KE and DE must determine whether the prototype is to be accepted for the next stage of development. This decision is not intended to be the end of the knowledge engineering, or even prototyping, process.

If the prototype is not sufficiently validated for prototype acceptance, **Further evaluation** is one option for the KE and DE. It will often be necessary to use multiple evaluation techniques to validate the model: for example, sensitivity to findings and parameter analyses evaluate different aspects of the model with little overlap, and hence don't substitute for each other. If problems with either the structure or the parameters have been identified, it will be necessary to re-visit the relevant KE processes, **Structural Development & Evaluation** or **Parameter Estimation** respectively, via the main spiral iteration in Figure 2.

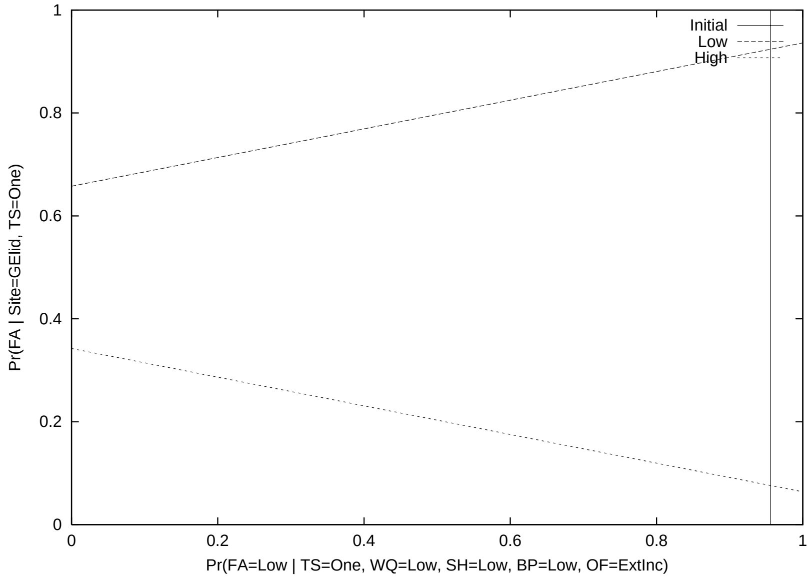

Figure 5: Linear functional representation of parameter sensitivity. Y-axis is the posterior probability, X-axis is the particular parameter (CPT entry).

6 Conclusion

In this paper we propose a methodology for combining expert information and data in parameterising BNs, an important research topic that has been widely acknowledged in the BN field but little developed. In many real world applications, including our ERA case study, one or both sources can be considered to be incomplete. Using the ERA case study, we developed and trialled a quantitative knowledge engineering methodology for the combination of such sources. The methodology has two major components, parameter estimation and quantitative evaluation. Incorporating the methodology into a BN knowledge engineering spiral, we developed a prototype that has been accepted for use in the ERA domain. Future studies will investigate using this methodology in other applications, and to iteratively assess and develop the deployed ERA model.

References

[1] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematics Society*, 35:99–110, 1943.

[2] B. W. Boehm. A spiral model of software development and enhancement. *IEEE Computer*, pages 61–72, 1988.

[3] T. Boneh. Visual investigation of graphical structure in Bayesian networks. Master's thesis, Department of Computer Science, University of Melbourne, 2003.

[4] F. Brooks. *The Mythical Man-Month: Essays on Software Engineering.* Addison-Wesley, Reading, MA, second edition, 1995.

[5] P. L. Cadwallader. Some causes of the decline in range and abundance of native fish in the Murray-Darling River system. In *The Murray-Darling River System Symposium, Proceedings of the Royal Society of Victoria*, 1978.

[6] Hei Chan and Adman Darwiche. When do numbers really matter? *Journal of Artificial Intelligence Research*, 17:265–287, 2003.

[7] V. M. H. Coupe and L. C. van der Gaag. Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36:323–356, 2002.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39:1–38, 1977.

[9] M.J. Druzdzel and L.C. van der Gaag. Building probabilistic networks: Where do the numbers come from? *IEEE Trans. on Knowledge and Data Engineering*, 12(4):481–486, 2001.

[10] W. D. Erskine. Downstream hydrogeomorphic impacts of Eildon Reservoir on the mid-Goulburn River, Victoria. *Proceedings of the Royal Society of Victoria.*, 108:1–15, 1996.

[11] C. J. Gippel and B. L. Finlayson. Downstream environmental impacts of regulation of the Goulburn River, Victoria. In *Hydrology and Water Resources Symposium, Newcastle*, 1993.

[12] D. Heckerman and D. Geiger. Learning Bayesian networks. In P. Besnard and S. Hanks, editors, *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (UAI95)*, pages 274–284, San Francisco, 1995.

[13] M. Henrion, M. Pradhan, B. Del Favero, K. Huang, G. Provan, and P. O'Rorke. Why is diagnosis using belief networks insensitive to imprecision in probabilities? In *UAI96*, pages 307–314, 1996.

[14] L. Hope, A.E. Nicholson, and K.B. Korb. Knowledge engineering tools for probability elicitation. Technical report 2002/111, School of Computer Science and Software Engineering, Monash University, 2002.

[15] O. Kipersztok and H. Wang. Another look at sensitivity of Bayesian networks to imprecise probabilities. In *Proceedings of the Eighth Workshop on Artificial Intelligence and Statistics (AISTAT-2001)*, Florida, 2001.

[16] K. B. Korb and A. E. Nicholson. *Bayesian Artificial Intelligence.* Computer Science and Data Analysis. CRC, Boca Raton, 2004.

[17] K.B. Laskey and S.M. Mahoney. Network engineering for agile belief network models. *IEEE: Transactions on Knowledge and Data Engineering*, 12:487–98, 2000.

[18] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. In G. Shafer and J. Pearl, editors, *Readings in Uncertain Reasoning*, pages 415–48. Morgan Kaufmann, 1990.

[19] A. Nicholson, T. Boneh, T. Wilkin, K. Stacey, L. Sonenberg, and V. Steinle. A case study in knowledge discovery and elicitation in an intelligent tutoring application. In Breese and Koller, editors, *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI01)*, pages 386–394, 2001.

[20] Norsys. Netica. http://www.norsys.com, 2000.

[21] A. Onisko, M.J. Druzdzel, and H. Wasyluk. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. In *Working Notes of the Workshop on "Bayesian and Causal networks: from inference to data mining." 12th European Conference on Artificial intelligence (ECAI-2000)*, 2000.

[22] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, CA, 1988.

[23] C.A. Pollino, P. Feehan, M. Grace, and B. Hart. Fish communities and habitat changes in the highly modified goulburn catchment. *AMarine and Freshwater Research (accepted)*, 2005.

[24] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, and K. Huang. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence*, 85(1-2):363–397, 1996.

[25] B. Rieman, J. T. Peterson, J. Clayton, P. Howell, R. Thurow, W. Thompson, and D. Lee. Evaluation of potential effects of federal land management alternatives. *Forest Ecology and Management*, 153(1-3):43–62, 2001.

[26] L. C. van der Gaag and S. Renooij. Analysing sensitivity data from probabilistic networks. In Breese and Koller, editors, *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI01)*, pages 530–537, San Franscisco, CA., 2001. Morgan Kaufmann.

[27] C.S. Wallace and K.B. Korb. Learning linear causal models by MML sampling. In A. Gammerman, editor, *Causal Models and Intelligent Data Management*. Springer-Verlag, 1999.