Operationalisation of natural capital and ecosystem services: from concepts to real-world applications

WP3 Methodological Guidelines for Bayesian Belief Networks

Ron Smith, Anders L Madsen, Roy Haines-Young, David Barton

Main contact: Ron Smith, NERC-CEH, ris@ceh.ac.uk

Table of contents

1. Types of problem the method can be used to study

A Bayesian Belief Network (BBN) is a framework that uses a graphical representation to show the flow of information in a system. It has **nodes** or **vertices** to represent variables (including observed quantities, latent (unobserved) quantities, expert opinion, model outputs, unknown parameters, ...) and **links** or **edges** joining ”parent” nodes to ”child” nodes (Figure 1a). The difference between this and other similar frameworks is in the use of conditional probabilities to express the relationships between nodes. This allows the building of complex networks from simple segments (Figure 1b) and it enables uncertainties to be assessed at every stage, so the outcomes of the network reflect the weight of the evidence that supports the conclusion.

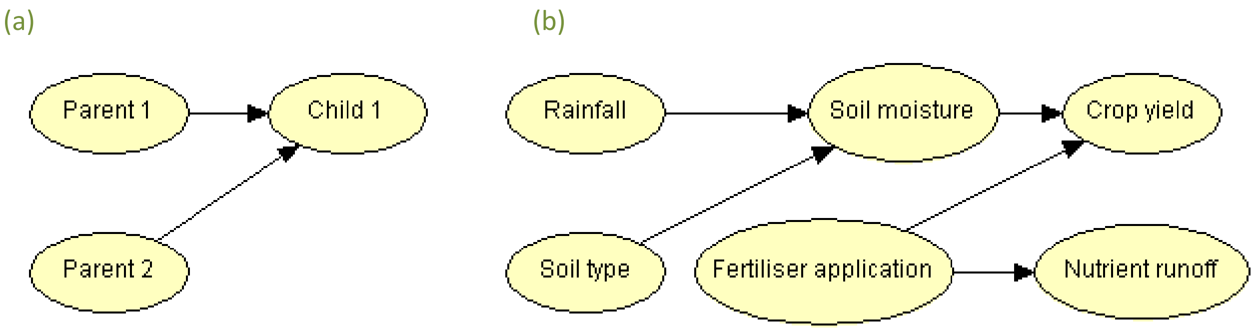

Figure 1: Examples of simple BBNs showing (a) the basic elements and (b) starting to build a more complex structure.

A BBN can be used for a variety of different problems. It can represent the chains of inputs and interactions leading to, for example, a land management decision, and so a BBN can be used as a tool in a decision or decision support context. The basic structure linking together models, data of different types and expert opinion, can also provide a simple meta-modelling and meta-analysis framework, with the latter particularly useful in interdisciplinary projects. In an ecosystem services context, BBNs can usefully represent one part of the decision process. They can potentially provide a linking structure from the models of ecological function, through delivery of service, social and/or economic assessment including regulatory (e.g. planning) constraints, and decision criteria such as trade-offs, cost-benefit or multi-criteria analyses. This latter element may require using Influence Diagrams (ID), an extension of BBNs that includes extra decision-making functionality within the framework. The facility of using a variety of information types including expert opinion helps with the rapid development and testing of a basic ecosystem services model when data may be scarce.

Some of the other benefits of using a BBN are that:

- it can learn from new data by updating its probabilities and so it always reflects the current state of knowledge,
- the possibility of achieving a chosen outcome can be easily assessed without having to set up multiple simulations, and
- the graphical interface helps to focus ideas during network development and encourages transparency about the system structure.

The interface also provides an attractive way of presenting the arguments to stakeholders. The networks are very good for quickly exploring options and building an understanding of the potential outcomes of scenarios, particularly useful within a participatory setting.

2. Data requirements

The data requirements are basically linked to the structure and to the purpose of the BBN. All types of data can be used for developing a BBN, but internally the nodes will have a discrete set of states and the information about the links is held as a set of conditional probability tables (CPTs). For computational purposes, it is a good idea to limit the number of states in any node and to limit the number of links going into any one node.

The BBN structure can be set up purely from expert opinion, and that is probably the most likely route for most case studies. Most networks are then developed iteratively, and if the nodes change during these iterative steps then so will the data requirement. With this approach, it is unlikely in an ecosystem services application that all data will or needs to be available for the initial version of a BBN. However, if a large dataset exists, it is also possible to derive the network structure as part of a data-mining exercise and this may be an alternative for some case studies.

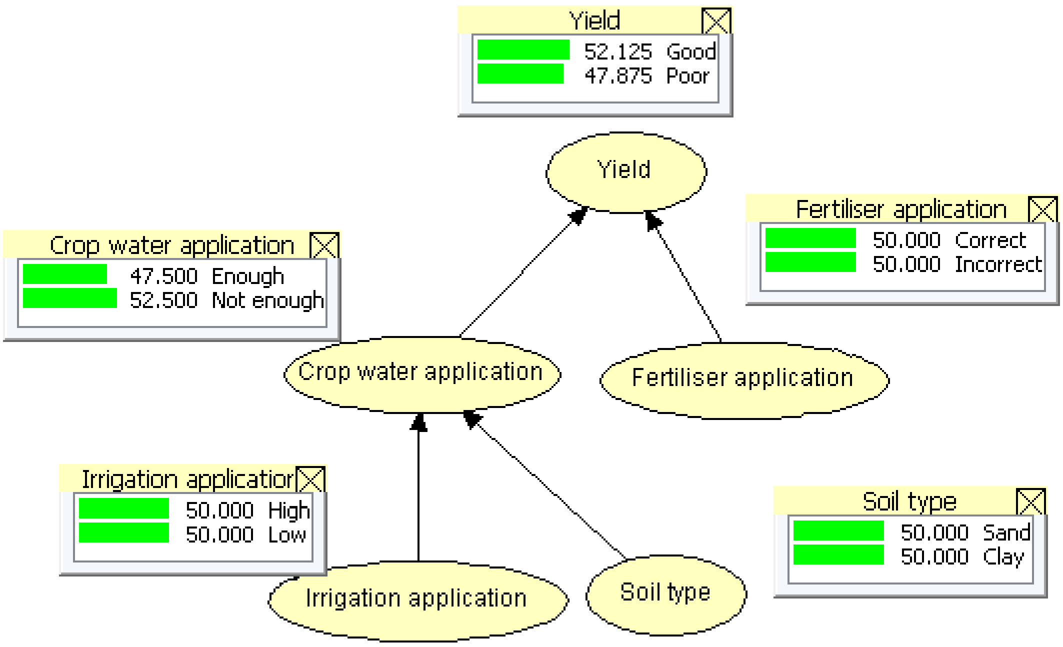

Figure 2: Example BBN to predict yield given management inputs (modified from Cain, 2001).

Figure 2 shows a very simple BBN with all nodes having only two states. The two CPTs for this BBN, Crop water application and Yield, are shown in Figure 3. The first column of the Yield CPT in Figure 3 shows the conditional probabilities assuming the fertiliser application is ‘correct’. This shows that in this situation the probability of going from ‘enough’ water to ‘good’ yield is 0.8 and of going from ‘enough’ water to ‘poor’ yield is 0.2. If the assumption is that the fertiliser application is ‘incorrect’, then the conditional probabilities change and now the probability of going from ‘enough’ water to ‘good’ yield is 0.4 and of going from ‘enough’ water to ‘poor’ yield is 0.6. These sets of conditional probabilities must add up to 1 to specify the BBN correctly.

Yield

Fertiliser application	Correct		Incorrect	
Crop water application	Enough	Not enough	Enough	Not enough
Good	0.8	0.6	0.4	0.3
Poor	0.2	0.4	0.6	0.7

Crop water application

Soil type	Sand		Clay	
Irrigation application	High	Low	High	Low
Enough	0.7	0.1	0.8	0.3
Not enough	0.3	0.9	0.2	0.7

Figure 3: Conditional Probability Tables (CPTs) defining the links in the yield model (Figure 2).

Information can be entered directly into the CPTs, so that it is pre-processed into the probabilities of going from each state (A,B,C,...) in parent node X to the various states (D,E,F,...) in child node Y. The source of these data in the above example could be a summary of published literature, a history of yields on a particular farm, or the ‘expert’ opinion of the farmer or a combination of these sources. For a larger BBN it is probably easier to generate these tables outside the BBN software and then either copy and paste them in or make a link directly to an Excel spreadsheet.

Information can also be entered into the CPTs as equations with the uncertainty specified as part of the input. The equation could be derived from the output of a regression model of the variable in the child node dependent on the parent node variables, and where the parameters and the estimate of residual variance from the regression become the inputs to the BBN. Again the data are being pre-processed before the information is entered into the BBN. If the relationship is taken from the literature, then some estimate of the variance will have to be used and the sensitivity of the BBN output to that value can be tested.

Data, in the form of cases, can also be entered directly into the BBN. Generally a file of cases is prepared where each case consists of a simultaneous set of observations of the states of some or all of the nodes in the BBN (Figure 4). These cases can then be used to update the CPTs (often referred to as learning from data) by combining this new data with the information already in the CPTs. If the CPTs were originally set to default (non-informative) values, then this process can be used to estimate the initial parameters in the CPT tables.

If the BBN is purely being used for scenario testing and the expert opinion or existing model outputs (including regression fits, etc.) can be substantiated outside the BBN, then no further data may be required. An example of this is where a link between amount of rainfall and water level in a river is already well established and further data about this link will not substantially improve the BBN. However the BBN will require information on the certainty/uncertainty of that relationship.

C1	C2	C3	C4	C5
Good	Not enough	Correct	High	Sand
Poor	Enough	Incorrect	High	Clay
Poor	Enough	Incorrect	Low	Clay
Poor	Not enough	Correct	High	Sand
Good	Enough	Correct	Low	Sand

Figure 4: Example of a case file relating to the Yield example. The headers C1 to C5 refer to the node names in the BBN, which in this example are Yield, Crop water application, Fertiliser application, Irrigation application and Soil type. Each subsequent row is a case where the data are the states of each node for that instance of the BBN – each instance could be a separate farm field, or the same field in a different year, for example. For input to Hugin this Excel file would be saved in CSV format.

3. Constraints and limitations

The major constraint on any ecosystem services assessment is the availability of information (data, model simulations, etc.) at the correct spatial and temporal scales, but the advantage of the BBN is that it can coherently combine expert opinion with other data so reducing the problems caused by missing information.

The nodes in Hugin (and similar software implementations of BBNs) usually operate with categorised rather than continuous variables. This is not normally a limitation to developing an effective network and there are functions in the software to convert continuous measurements to the chosen discrete states. The random variables (nodes) have to be precisely defined and have states that are finite, exhaustive, and mutually exclusive. For each parent configuration a probability distribution over the child states has to be provided, so all the conditional probabilities have to sum to 1.

In all ecosystem services assessments decisions are made about the boundaries to the problem, e.g. spatial extent of the ecosystem or time period for the assessment. The BBN construction makes a further constraint more obvious, in that there is a decision to be made about the complexity of the network to be modelled. For computational reasons, it is better to keep to a relatively few number of states in any node (i.e. 1-10 rather than 100s) and to keep down the number of parents for each child node (i.e. less than 10), but intermediate nodes and the use of subnets, for example, may be used to ensure there is no loss of important information.

Technically a BBN is one of a class of directed, acyclic graphs and the last part of that definition means that there can be no closed loops in the diagram, so no node can influence itself. This precludes there being feedbacks within any one time instance of the BBN, but the use of time slicing (i.e repeating the BBN at a series of points in time) can overcome some of this problem. An ecosystem services study is specific to a defined time period, and careful choice of period can ensure any feedback does not occur within any one time slice.

4. Steps required to apply the method within a case study

The first task is to define the structure using the graphical display. At the initial stages this can be a good way of thinking through how the system works, whether this be components of an ecological function model or a socio-economic decision tool. One important consideration is the level of detail required for the various elements of your network and this can be explored as the network develops, which is usually an iterative process involving tests on the network (e.g. sensitivity analysis), comparisons of data with prior assumptions, and stakeholder consultation.

It can be useful to involve stakeholders early on in the process, especially if the BBN will be used in a decision support mode. If the BBN is being used as a tool for meta-modelling or meta-analysis then there maybe be no stakeholders except for the research teams. Depending on the level of initial knowledge, the stakeholders may derive narratives in the first instance or, where this is possible, they may go straight to producing a graphical diagram of part or all of the network. The diagram should be comprehensible and it's better that the nodes have a readily meaningful interpretation. This part of the stakeholder consultation may be integrated into the more general stakeholder consultations for the project.

The next stage (Figure 5) is to develop the conditional probability tables (CPT) for each node. These are conditional probabilities because they express the probabilities of the states in the child node conditional on the values in the states of the parent nodes. These may be derived from the literature, from data, or from expert opinion, or a combination. Data used at this juncture may be analysed separately from the BBN and the results entered as a set of probabilities. If stakeholders are providing the expert opinion, it may be worth looking at specialist elicitation software such as the MATCH Uncertainty Elicitation Tool: it can be used for free and is available at http://www.match.ac.uk/uncertainty/ (Morris et al, 2013).

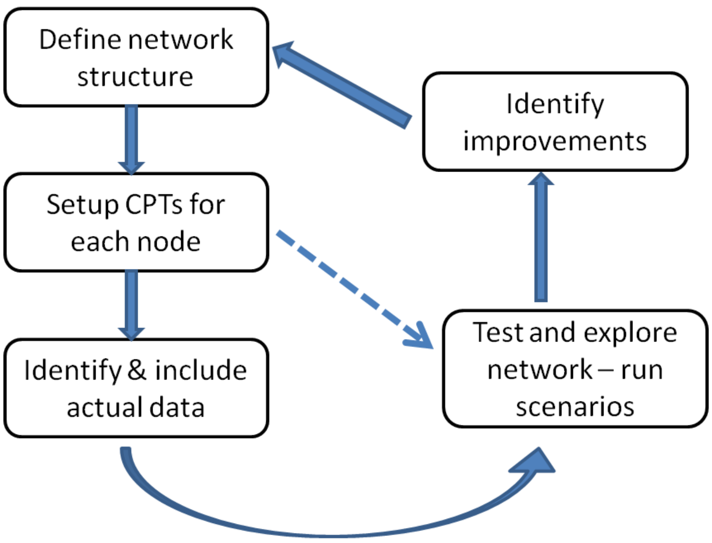

Figure 5: Flow diagram of the stages.

If actual data are available for the study, these are introduced at this point so the BBN can learn from this additional information. These data are input to the BBN through case files (see above) and may be for all the nodes in the network or only for part of the network. The use of conditional probabilities within a BBN allows the network to accommodate missing values in the case files, and nodes with latent variables, by definition, will not have any observations. The spatial extent of the BBN will determine how geographical data are processed. In the crop yield example, is the decision process applied at a field level or at a farm level? Are there trade-offs between fields on a farm? The answers to questions of this type will determine what spatial and temporal averaging is required for the data.

Testing and exploring the network should follow. This may involve introducing scenario data and seeing how the network performs, or using some sensitivity tests of particular nodes or parts of the network. If the network is getting too complicated, it may be helpful to isolate parts of the BBN into subnets with, in Hugin, an instance node retained in the main BBN to represent the outcome of the subnet.

Most complex network developments will involve iteration back through all these stages, possibly several times, before both developers and stakeholders are satisfied that the network adequately represents the problem and there are sufficient data for the network to produce useful outputs.

5. Example applications

The OpenNESS Hugin website (http://openness.hugin.com/) is collecting examples of applications of BBNs within the project, and provides a good opportunity to see some BBNs in action. Some examples from the literature are included in Further Reading below.

6. Further reading

See also "Bayesian Belief Networks: A Cross-Cutting methodology in OpenNESS: Briefing Note" from the OpenNESS website.

The in-depth introduction for Hugin is in the book:
Uffe B. Kjærulff and Anders L. Madsen (2013). Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. Springer. ISBN: 978-1-4614-5103-7 (Print) 978-1-4614-5104-4 (Online).

There are a lot of papers now about applying Bayesian belief networks in ecology, and some are listed below:

Amstrup, S.C., Marcot, B.G. and Douglas, D.C. (2008). A Bayesian network modeling approach to forecasting the 21st century worldwide status of polar bears. Geophysical Monograph 180: 213-268.

Celio, E., Koellner, T. And Gret-Regamey, A. (2014). Modelling land use decisions with Bayesian networks: Spatially explicit analysis of driving forces on land use change. Environmental Modelling and Software, 52: 222-233.

Chen, S.H. and Pollino, C.A. (2012). Good practice in Bayesian network modelling. Environmental Modelling and Software, 37: 134-143.

Landuyt, D., Broekx, S., D'hondt, R., Engelen, G. And Aertsens, J. (2013). A review of Bayesian belief networks in ecosystem service modelling. Environmental Modelling and Software, 46: 1-11.

Pitchforth, J. and Mangerson, K. (2013). A proposed validation framework for expert elicited Bayesian Networks. Expert Systems with Applications, 40: 162-167.

Spence, P.L. and Jordan, S.J. (2013). Effects of nitrogen inputs on freshwater wetland ecosystem services – A Bayesian network analysis. Journal of Environmental Management, 124: 91-99.

Sun, Z. And Muller, D. (2013). A framework for modelling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. Environmental Modelling and Software, 45: 15-28.

Troldberg, M., Aalders, I., Towers, W., Hallett, P.D., McKenzie, B.M., Bengough, A.G., Lilly, A., Ball, B.C. and Hough, R.L. (2013). Application of Bayesian Belief Networks to quantify and map areas at risk to soil threats: Using soil compaction as an example. Soil and Tillage Research, 132: 56-68.

7. References

Morris, D.E., Oakley, J.E. and Crowe, J.A. (2013). A web-based tool for eliciting probability distributions from experts. Environmental Modelling and Software, 52: 1-4. DOI: 10.1016/j.envsoft.2013.10.010.