Bayesian Network Training Workshop

A practical workshop on the development of Bayesian Networks for Regional Natural Resource Managers of Australia.

Conducted by the iCAM Centre,
The Fenner School for Environment and Society
Australian National University, Canberra

Sponsored by Landscape Logic,
a Commonwealth Environment Research Facilities project

© Integrated Catchment Assessment and Management (iCAM),
The Australian National University, 2007
Workshop Objectives

- Introduce Bayesian Networks (BNs) as integration tools
- Present the process for building BNs:
  - Conceptual models
  - Influence diagrams
  - Inputs into BNs
  - Evaluating BNs
- Undertake workshop exercises to build preliminary BN models
### Agenda: Day 1

<table>
<thead>
<tr>
<th>9.00 am – 9.15 am</th>
<th>Introductions</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.15 am – 10.15 am</td>
<td>Training workshop outline</td>
</tr>
<tr>
<td><strong>MODULE 1. Introduction to BNs</strong></td>
<td></td>
</tr>
<tr>
<td>o Building BNs</td>
<td></td>
</tr>
<tr>
<td>o Software for BNs</td>
<td></td>
</tr>
<tr>
<td><strong>Morning Tea</strong></td>
<td></td>
</tr>
<tr>
<td>10.30 am – 12.30 pm</td>
<td>MODULE 2. Conceptual models and influence diagrams</td>
</tr>
<tr>
<td>o Defining your asset/value/scope</td>
<td></td>
</tr>
<tr>
<td>o Identify threats and management activities</td>
<td></td>
</tr>
<tr>
<td>o Building a conceptual model</td>
<td></td>
</tr>
<tr>
<td>Workshop exercise: Develop conceptual model</td>
<td></td>
</tr>
<tr>
<td>o Influence diagrams to BN</td>
<td></td>
</tr>
<tr>
<td>o Netica demonstration</td>
<td></td>
</tr>
<tr>
<td>Workshop exercise: Transform conceptual model into influence diagram in Netica</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12.30 pm – 1.15 pm</th>
<th>Lunch</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.15 pm – 3.00 pm</td>
<td>Workshop exercise: Transform conceptual model into influence diagram in Netica (cont.)</td>
</tr>
<tr>
<td><strong>MODULE 3. Defining variable states</strong></td>
<td></td>
</tr>
<tr>
<td>o Sources of information</td>
<td></td>
</tr>
<tr>
<td>o Qualitative</td>
<td></td>
</tr>
<tr>
<td>o Quantitative</td>
<td></td>
</tr>
<tr>
<td>Workshop exercise: Defining states</td>
<td></td>
</tr>
<tr>
<td><strong>Afternoon Tea</strong></td>
<td></td>
</tr>
<tr>
<td>3.15 pm – 4.15 pm</td>
<td>Workshop exercise: Defining states (cont.)</td>
</tr>
<tr>
<td><strong>MODULE 4. Condition Probability Tables</strong></td>
<td></td>
</tr>
<tr>
<td>o Sources of information for models</td>
<td></td>
</tr>
<tr>
<td>o Input of expert elicitation into BNs</td>
<td></td>
</tr>
<tr>
<td>Basics</td>
<td></td>
</tr>
<tr>
<td>Difference of opinion</td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td></td>
</tr>
<tr>
<td>4.15 pm – 4.30 pm</td>
<td>Feedback/review discussion</td>
</tr>
</tbody>
</table>

### Agenda: Day 2

<table>
<thead>
<tr>
<th>9.00 am – 9.15 am</th>
<th>Overview of Day 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plan for Day 2</td>
</tr>
<tr>
<td>9.15 am – 10.15 am</td>
<td>Workshop exercise: Expert elicitation</td>
</tr>
<tr>
<td><strong>Morning Tea</strong></td>
<td></td>
</tr>
<tr>
<td>10.30 am – 11.00 am</td>
<td>MODULE 4. Condition Probability Tables (cont.)</td>
</tr>
<tr>
<td>o Input of quantitative data into BNs</td>
<td></td>
</tr>
<tr>
<td>o Using monitoring data</td>
<td></td>
</tr>
<tr>
<td>o Using existing models</td>
<td></td>
</tr>
<tr>
<td>o Spatial data</td>
<td></td>
</tr>
<tr>
<td>o Temporal data</td>
<td></td>
</tr>
<tr>
<td>o Examples</td>
<td></td>
</tr>
<tr>
<td>o Strengths and weaknesses of BNs</td>
<td></td>
</tr>
<tr>
<td>11.00 am – 11.30 am</td>
<td>Workshop exercise: Identifying potential data sources for your BN</td>
</tr>
<tr>
<td>11.30 am – 12.15 pm</td>
<td><strong>MODULE 5. Model Evaluation tools</strong></td>
</tr>
<tr>
<td>o Why? How? What it means?</td>
<td></td>
</tr>
<tr>
<td>o Sensitivity analysis</td>
<td></td>
</tr>
<tr>
<td>o Knowledge gaps</td>
<td></td>
</tr>
<tr>
<td>o Predictive accuracy</td>
<td></td>
</tr>
<tr>
<td>12.15 pm – 1.00 pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>1.00 pm – 2.30 pm</td>
<td>Workshop exercise: Finalise Draft BNs</td>
</tr>
<tr>
<td>2.30 pm – 3.15 pm</td>
<td>o Feedback</td>
</tr>
<tr>
<td>o Where to from here?</td>
<td></td>
</tr>
<tr>
<td>o Workshop Evaluation</td>
<td></td>
</tr>
<tr>
<td><strong>Afternoon Tea</strong></td>
<td></td>
</tr>
</tbody>
</table>
Module 1: Introduction to Bayesian Networks

Building BNs

Software for BNs

Issues of NRM Sustainability

Land and Water Resources

Urban Development

Recreation & Tourism

Cultural Heritage
  Indigenous & European

Social Amenity

Environmental Conservation
  Terrestrial & Aquatic Ecology

Agriculture

Complex interactions where models and decision support systems can assist
NRM Models

- In NRM, we need models that:
  - Inform decision making
  - Represent complex systems (i.e. multiple interacting variables)
  - Incorporate knowledge and data
  - Quantify uncertainties
  - Promote adaptive management
  - Scientifically robust

Integration Approaches

<table>
<thead>
<tr>
<th>Model Purpose</th>
<th>System dynamics</th>
<th>Bayesian networks</th>
<th>Meta models</th>
<th>Coupled complex models</th>
<th>Agent based models</th>
<th>Expert systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Forecasting</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Decision making</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>System understanding</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Social learning</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Qualitative and quantitative</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Quantitative only</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Focused and indepth</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>General and broad</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Compromise</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Both</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input Data Type</th>
<th>Qualitative and quantitative</th>
<th>Quantitative only</th>
<th>Focused and indepth</th>
<th>General and broad</th>
<th>Compromise</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal Range</td>
<td>Qualitative and quantitative</td>
<td>Quantitative only</td>
<td>Focused and indepth</td>
<td>General and broad</td>
<td>Compromise</td>
<td>Both</td>
</tr>
<tr>
<td>Express uncertainty</td>
<td>Yes</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Model Output</td>
<td>Individual</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td>Aggregated</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
</tbody>
</table>
Bayesian Network Strengths

- Model complex systems
- Integrate across disciplines
  - Hydrology, water quality, ecology
  - Economics, social, environmental
- Integrate different information sources to utilise the best information available
  - Expert assessments, monitoring data, simulation models, research data
- Take into account uncertainties
- Guide data collection
  - Targeted to benefit the understanding (and management) of a system
- Models iteratively updated (adaptive management)
- Development process engages multiple people/groups (not an individual activity)

BNs and Scenarios

- Scenarios – explore ‘What if?’
- Test management or system changes
  - Examine relative change of probabilities which conveys expected system response while accounting for predictive uncertainties
- Therefore can assist in planning and investment decisions
LL: Why BNs?

- Relatively easy to represent complex systems so its comparatively easy to learn the process.
- Process promotes system learning
- Easy to integrate quantitative and qualitative information
- Utilise best information available, and can be easily updated
- Represent uncertainty
- Can document current assumptions used for making decisions
- Can be used to identify key knowledge gaps

Why is Uncertainty so Important?
Uncertainty & NRM

• Major sources of uncertainty:
  – Analytical, due to limited information and observation error (incomplete knowledge - reducible)
  – Variability and randomness of the system (not reducible)

• Natural resource Management
  – Quantification of uncertainty important → describe realistic outcomes and add flexibility to the decision process
Acknowledging uncertainty – small risk

Optimal yield → Environmental harm

little → Fertiliser application → lots

Acknowledging uncertainty – large risk

Optimal yield → Environmental harm

little → Fertiliser application → lots
Reflection of Uncertainty

High degree of certainty

<table>
<thead>
<tr>
<th>Output states</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10mg/L</td>
<td>0</td>
</tr>
<tr>
<td>10-20mg/L</td>
<td>0.2</td>
</tr>
<tr>
<td>20-40mg/L</td>
<td>0.6</td>
</tr>
<tr>
<td>&gt;40mg/L</td>
<td>1</td>
</tr>
</tbody>
</table>

Low degree of certainty

<table>
<thead>
<tr>
<th>Output states</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10mg/L</td>
<td>0.2</td>
</tr>
<tr>
<td>10-20mg/L</td>
<td>0.4</td>
</tr>
<tr>
<td>20-40mg/L</td>
<td>0.6</td>
</tr>
<tr>
<td>&gt;40mg/L</td>
<td>0</td>
</tr>
</tbody>
</table>

NOTE: If states are coarser, there is a higher perceived certainty in the output

Utilising Available Information

- We still make decisions when knowledge is poor
  - Capture assumptions and embed them in a framework where these can be tested
- Can test the sensitivity of assumptions
- Once assumptions are defined within the framework, management decisions are repeatable
- Increases transparency of management decisions
Probability Definitions

1. A probability $P(A)$ is a number between 0 and 1

2. Joint probability is the likelihood that condition A and condition B will occur.
   \[ P(AB) = P(A) \times P(B) \]

3. Conditional probability is the likelihood of A being in a condition, given the condition of B.
   \[ P(A|B) = \frac{P(AB)}{P(B)} \]

Use of a BN

- Probabilities tell us about the likelihood of an event occurring and the consequences of this upon other variables
- We can use Bayes’ Theorem to investigate the likelihood and test it with new data
- We can use BNs to update and improve our knowledge of a system as new information/evidence is acquired
BNS - Definitions

• Prior probability
  – Likelihood the variable will be in a particular condition given the current understanding and assumptions

• Posterior probability
  – The likelihood that a variable will be in a particular condition given the input variables, the conditional probabilities, and rules governing how probabilities combine

Bayes’ Theorem

\[ P(H | E) = \frac{P(E | H) \cdot P(H)}{P(E)} \]

Where \( H = \text{hypothesis} \) and \( E = \text{evidence} \)

Rev. Thomas Bayes (1702-1761)
Example

• After increased learning about how lake nitrogen impacts upon Algal Blooms

• \( P(A|TN) = \frac{P(TN|A) \times P(A)}{P(TN)} \)

Let \( A = \) algal bloom, \( TN = \) lake nitrogen
\( P(A|TN) = \) Probability of algal bloom given the nitrogen
\( P(TN|A) = \) Probability of nitrogen given algal bloom
\( P(A) = \) Probability of algal blooms

Bayesian Networks

*Have 2 components: Structure and Probabilities*

Observed data
Model simulation
Expert opinion
Literature

Outputs can be qualitative or quantitative
e.g. Decrease, No Change, Increase
e.g. >10 ha decrease, <10 ha decrease, No change, <10 ha increase, >10 ha increase
Example

### Decision: Foreshore vegetation
- Scenario options = 2
  1. No change
  2. 50m buffer

### Decision: Urban development
- Scenario options = 3
  1. No change
  2. Low density
  3. Medium density

- **Variable:** Visual amenity
  - Input link = 1
  - States = 3
    1. Decrease
    2. No change
    3. Increase

- **Variable:** Lake nitrogen
  - Input link = 2
  - States = 2
    1. < 20mg/L
    2. >20mg/L
Example

Variable:
Recreation
Input link = 2
1. Visual amenity
2. Lake nitrogen
States = 3
1. Decrease
2. No change
3. Increase

Variable:
Seagrass
Input link = 1
1. Lake nitrogen
States = 3
1. Decrease
2. No change
3. Increase

Utility:
Recreational revenue
Input link = 1
1. recreation
Model output = single
$ value

Causal Links

• The relationship between nodes joined by a link is described by a probability distribution.

For example:
• Impact upon Visual amenity is determined by the scenario option chosen for foreshore vegetation.
• Each scenario option can generate a different probability distribution
For each Foreshore vegetation scenario option there is a predetermined impact (conditional probability) upon Visual amenity, defined by its three output states.

**No change**
- Probability Visual amenity
  - Decreases = 0.15
  - No change = 0.75
  - Increases = 0.10

**50m buffer**
- Probability Visual amenity
  - Decreases = 0
  - No change = 0.25
  - Increases = 0.75

**Question:** What is the predetermined impact upon Recreation given the possible output states of Visual amenity?

- **Decrease in Visual Amenity**
  - Probability Recreation
    - Decreases = 0.6
    - No change = 0.4
    - Increases = 0

- **No change in Visual Amenity**
  - Probability Recreation
    - Decreases = 0.05
    - No change = 0.90
    - Increases = 0.05

- **Increase in Visual Amenity**
  - Probability Recreation
    - Decreases = 0
    - No change = 0.4
    - Increases = 0.6
**Variable to Utility**

- **Question:** What is the expected change in value of Recreational revenue given change in recreation?

Value assigned to each output state of Recreation:
- Decrease = -$1000
- No change = $0
- Increase = $2000

Recreation output distribution = A
Output value equals sum of Recreation:
- Decreasing = 0.6 * -1000
- No change = 0.4 * 0
- Increasing = 0 * 2000
TOTAL = -$600

Recreation output distribution = B
Output value equals sum of Recreation:
- Decreasing = 0.05 * -1000
- No change = 0.9 * 0
- Increasing = 0.05 * 2000
TOTAL = $50

---

**Decision to Variable to Variable**

If scenario option 2 (50m buffer) was selected, the probability of:
- Decrease in Visual amenity = 0
- No change in Visual amenity = 0.25
- Increase in Visual amenity = 0.75

If Visual amenity changes as follows, the probability for the output states is:
- Decreases = 0.60, 0.40, 0
- No change = 0.05, 0.90, 0.05
- Increases: = 0, 0.40, 0.60
The final single probability distribution for Recreation equals the sum of the probability of Visual amenity \(p(\text{Vis})\) having each of its output states, multiplied by the outcome if that occurs (Joint probability).

\[
\begin{array}{c|ccc}
\text{p(Vis)} & \text{p(RDe)} & \text{p(RNC)} & \text{p(Rin)} \\
0 & 0.60 & 0.40 & 0.00 \\
0.25 & 0.05 & 0.90 & 0.05 \\
0.75 & 0.00 & 0.40 & 0.60 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{p(E1)} & \text{p(E2)} & \text{p(E3)} \\
0 \times 0.6 = 0 & 0 \times 0.4 = 0 & 0 \times 0 = 0 \\
0.25 \times 0.01 = 0 & 0.25 \times 0.23 = 0.01 & 0.25 \times 0.01 = 0.01 \\
0.75 \times 0 = 0 & 0.75 \times 0.4 = 0.3 & 0.75 \times 0.6 = 0.45 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{p(Rin)} & \text{p(RNC)} & \text{p(RDe)} \\
0 & 0.40 & 0.00 \\
0.25 & 0.05 & 0.90 \\
0.75 & 0.00 & 0.40 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{p(E1)} & \text{p(E2)} & \text{p(E3)} \\
0.75 \times 0.6 = 0 & 0.75 \times 0.4 = 0 & 0.75 \times 0 = 0 \\
0.25 \times 0.05 = 0.01 & 0.25 \times 0.9 = 0.23 & 0.25 \times 0.01 = 0.01 \\
0.75 \times 0 = 0 & 0.75 \times 0.4 = 0.3 & 0.75 \times 0.6 = 0.45 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{p(Vis)} & \text{p(Rin)} & \text{p(RNC)} \\
0 & 0.75 & 0.25 \\
0.25 & 0.75 & 0.25 \\
0.75 & 0.25 & 0.00 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{p(E1)} & \text{p(E2)} & \text{p(E3)} \\
0 \times 0.6 = 0 & 0 \times 0.4 = 0 & 0 \times 0 = 0 \\
0.25 \times 0.01 = 0 & 0.25 \times 0.23 = 0.01 & 0.25 \times 0.01 = 0.01 \\
0.75 \times 0 = 0 & 0.75 \times 0.4 = 0.3 & 0.75 \times 0.6 = 0.45 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{p(Rin)} & \text{p(RDe)} & \text{p(RNC)} \\
0 & 0.40 & 0.00 \\
0.25 & 0.90 & 0.05 \\
0.75 & 0.25 & 0.00 \\
\end{array}
\]
Underlying Thinking of a BN

The Real World

Conceptual Model

Influence Diagram

Bayesian Network

Developing a Bayesian network

What is the objective of the model

Review literature, plans, etc.

Define asset, value to be considered

Conceptual model of how the system works

Link management actions and threats to assets and values

Transform conceptual model in influence diagram

Consultation within the regional group and external stakeholder (e.g. state govt agency staff)

Describe the model variables (states)

Population & document assumptions and information used in BNs

Evaluation models (sensitivity and accuracy)

Testing model scenarios

Updating models
How do I Build a BN?

- Chose software:
  - We are using Netica (www.norsys.com)
- Define your model scope
  - Spatial and temporal scales
- Specify your structure
  - Conceptual model: Expert Elicited
  - Automated (Structure learning)
- Specify probability distributions
  - Priors (flat, expert, other models or data)
  - Use inference algorithms (L&S, EM, GD) to update models

How do I Build a BN?

- Test model robustness
  - Predictive accuracy
  - Sensitivity analysis
  - Expert evaluations
- If necessary, revise model and Test
  - Structure and CPDs
- Scenario Analysis
- Thorough reporting of model is essential (how structure and CPDs were derived)
Choose a Software Package

<table>
<thead>
<tr>
<th>BN packages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• DeNA in ICMS</td>
</tr>
<tr>
<td>• Analytica</td>
</tr>
<tr>
<td>• BayesiaLab</td>
</tr>
<tr>
<td>• Bayes Net Toolbox</td>
</tr>
<tr>
<td>• Deal (package available in R)</td>
</tr>
<tr>
<td>• Genie</td>
</tr>
<tr>
<td>• Netica</td>
</tr>
<tr>
<td>• Hugin</td>
</tr>
<tr>
<td>• WinBUGs</td>
</tr>
</tbody>
</table>

http://www.cs.ubc.ca/~murphyk/Software/BNT/bnsoft.html

Notes..............................................
Module 2. Conceptual Models and Influence Diagrams

Define Model Scope

• Define model objective (what is the question?)
• Define model scales (what are the bounds?)
  – Space and Time
• Important points:
  – More complex model: ↑ parameters, ↑ level of uncertainty (more parameters need to be estimated)
  – Matter of balance: holistic (general understanding) vs. reductionalist (fine detail)
    • This depends on your objective
  – BNs are not bound by lack of data and knowledge
    BUT the quality of your model is
Tip:

Keep the model as simple as possible and as complex as needed

Jørgensen and Bendoricchio

Specify Model Structure

Spatial and Temporal Representation in BNs
Scales in BNs

• Need to specify scales upfront
• Need to identify which variables / relationships change given a defined spatial and temporal scale
  – For example:
    • Given that all system drivers (e.g. temperature, light, nutrients, flow) are the same, will algal biomass be different at Site A compared to Site B?
    • Are the sources of domestic sewage at Site A different to those at Site B?
    • Are water quality conditions more influential in the dry season compared to the wet season? If so, why?
• If they are specific to a scale, they should be linked to scale variables using an arrow / arc

Spatial Data

• Needs to be pre-defined
  – Dependent on scale of interest (particularly for management)
  – Dependent on scale of data collection
• The finer spatial resolution, the greater the complexity
• Hierarchical representation in BNs
  – Catchment → Reach → Site
Data Across Spatial Scales

- Including spatial scales in your model will be dependent on (a) if it is important, and (b) what information you have
- Things you need to ask your groups:
  - What scale is important?
    - Sites
    - Reaches
    - Catchments
    - Countries
  - Do we have data across spatial scales?
    - YES → include spatial scales in your model
    - NO → you CANNOT include spatial scales in your model
Temporal Data

• Also needs to be pre-defined
  – Dependent on scale of interest (particularly for management)
  – Dependent on scale of data collection
• Models are NOT dynamic – they represent static predictions over defined temporal scales
• Finer temporal scales, greater complexity
• Dynamic models – Dynamic Bayesian Networks
Data Across Time

• Including time in your model will be dependent on (a) if it is important, and (b) what information you have

• Things you need to ask your groups:
  – What scale is important?
    • Annual predictions
    • Seasonal
    • Spans over different years (before versus after dam construction)
  – Do we have data across time?
    • YES → include time in your model
    • NO → you CANNOT include time in your model
Conceptual Models

- What are they?
- Why use them?
- How to build conceptual models

Based on information provided by Bill Dennison & Tim Carruthers (http://ian.umces.edu)

Conceptual Models

- A model of the system in which the main components and the processes connecting these components are explicitly shown
Why use Conceptual Models?

- Force the developers to identify the system they are working on
- Assist with clarifying current thinking / understanding
- More clearly identify the parts of the system the BN is focused upon
- Helps to ensure other stakeholders are all talking about the same system or parts of it
- Identify gaps/priorities/essential elements
- Communication
  (One way and two way - idea presentation and idea development)
Conceptual diagrams can capture the increasing understanding of a system...

Conceptual diagrams can depict processes at different scales...

© Integrated Catchment Assessment and Management (ICAM), The Australian National University, 2007
Types of Conceptual Diagrams

- Symbols (icons)

- Boxes and arrows / Spaghetti diagrams / Influence diagrams
WAVE-DOMINATED ESTUARY: NUTRIENT DYNAMICS

1) Nitrogen (both particulate and dissolved, or Total Nitrogen (TN)) enters the estuarine system from point- and non-point sources from within the catchment. River flow and nutrient input varies regionally, depending on local catchment and climatic conditions. However, the input of catchment-derived nutrients into estuaries and deltas is typically high (Harris, 2001).

2) Input of particulate N (PN) from atmospheric sources such as smoke and ash is significant in some wave-dominated estuaries.

3) The DIN is transported into the central basin of the estuary, with biological uptake by phytoplankton (Galloway et al., 1999; Webster et al., 2002), seagrass, macroalgae, and macrophytes occurring along the way, if residence times are long enough, and if temperature, turbidity, and light levels are suitable. The balance between planktonic and benthic primary productivity may depend on catchment nitrogen loads (Fry et al., 2002).

4) Some deposition and burial of particulate nitrogen (PN) occurs on flanking environments, due to the building effect of sediments and/or mangroves (where present) vegetation (Koon et al., 1988b; Nield et al., 1991; Howes et al., 1994; Langloisqvar, 1991). Burial and non-sorption of PN and dissolved inorganic nitrogen (DIN) can also occur within intertidal flats (Alongi et al., 1999). Some PN may be deposited and buried within the fluvial delta.

5) PN is deposited in the sediments as phytoplankton debris.

6) Decomposition of organic matter within the sediment produces dissolved inorganic nitrogen (potentially available for further plant/phytoplankton growth). Denitrification within the sediment converts nitrate (NO₃⁻) to N₂ gas, which escapes from the system to the atmosphere (Hogg et al., 1999). Some of the particulate nitrogen (PN) deposited into the sediment of the central basin is buried.

7) Seagrasses take up dissolved inorganic nitrogen (DIN) from the water column, and from the sediment pore-water (Hill et al., 1984; Hill et al., 1992). The pore-water dissolved inorganic nitrogen (DIN) is derived from the metabolism of phytoplankton, seagrass and other organic matter debris. The seagrass debris therefore, in part, is “recycled” back to the plants. N-fixation (incorporation of atmospheric N₂ to form nitrogenous organic compounds) occurring in the root zone contributes additional DIN to the root zone (Umbanhowar et al., 1985; Smiths, 1990). Denitrification is also an important process in seagrass meadows. Sandy sediments are permeable, hence can be ventilated by oxygen-rich overlying waters resulting in efficient remineralisation of organic debris (mostly by denitrification) with little preservation of organic matter.

8) Due to the long residence time typical of wave-dominated estuaries, most catchment-derived N is processed and effectively trapped by the estuary. Typically, only very small quantities of the TN load are exported to the marine environment, however export is much larger during significant during flood events (Nixson et al., 1996).

Spaghetti Diagram Example

- Planning
- Extension
- Reserves system
- Incentives
- Education

- Management of chemical use
- Pest management
- Fragmentation of native vegetation
- Pressures from wild flora harvesting
- Impacts of fire
- Condition of native vegetation
- Browsing by native animals
- Stock access
- Rural tree decline and death
- Disease management

Condition of Native Vegetation
Workshop Exercise

Conceptual Models
*Suggest to start with a symbol/pictorial conceptual model to assist in keeping it simple, concentrating on key aspects*

Conceptual Models to Influence Diagrams to Bayesian Networks
Conceptual Model Example

Influence Diagram Process

What is the issue (Asset/value)

Spatial scale
- Upstream/downstream
- Local/catchment-wide/off shore

Temporal scale
- Present
- Future (10 y, 30 y)

Influencing factors
(threats and management activities)

Influence diagram shows interconnections between assets/values, threats and activities
To Convert to an Influence Diagram

- Defined in conceptual model
  - asset /value
  - Influencing factors (threats and management activities)
- Now make the links

![Influence Diagram]

Influence Diagram

- Management Actions
- Riparian plantings
- Education
- Landuse: Cropping, Grazing, Urban development
- Runoff
- Sediment
- Nutrients
- Pesticides
- Water quality
- Water quantity
- Aquatic habitat
- Asset: River health
- Threats
Influence Diagrams to BN models

- Keep these points in mind to make the next step to BNs easier
- BNs can not have a feedback loop
  - Called a Directed Acyclic Graph (DAG)
- Criteria for inclusion of variables: Are they
  - Measurable,
  - Observable, or
  - Predictable
- We want no ambiguity in our definition of variables
  - Do we just look at ‘Fish’ in our models, or ‘Fish abundance’ / ‘Fish recruitment’ / ‘Fish biodiversity’?

BN Structure

- Goals in specifying BN structure
  - To minimise probability elicitation
    - Fewer nodes, fewer arcs, fewer states
  - Maximise ‘truthfulness’ of model
    - May requires more nodes, arcs, states
    - Balance benefit against cost of additional modeling
    - Too much detail can decrease accuracy
    - Too little detail can decrease model representativeness
- OCCAM’S RAZOR:
  - All things being equal, the simplest solution tends to be the best one
When faced with complex diagrams

• Use unobserved nodes for aggregating variables (i.e. develop sub-networks)
• This will reduce the number of conditional probabilities that need attained, simplifying the process
• As a general rule: Have no more than 4 parents per child variable
• Start thinking of simplifying the model structure early in the model development – don't wait until the parameterisation stage!

Black box Network (Complex Example)

482946 Conditional Probabilities
Tip:

Maintain system thinking – the model is NOT going to be a correct representation of reality, but an attempt to describe important system features

Jørgensen and Bendoricchio
Netica Practical Demonstration

Adding variables and links

Workshop Exercise

Using Netica (Group e.g.)
Conceptual models to Influence
Diagrams
Module 3. Defining Variable States

Sources of information

Qualitative
Quantitative

What are the States?

• Types of variable
  – Binary (2-valued)
  – Qualitative
  – Numeric discrete
  – Numeric continuous

• Dealing with infinite and continuous state sets
  – Some Bayes net software requires discretizing continuous variables
  – State ranges should represent meaningful differences in effect on related variables
The Clarity Test

- Where possible, final model structure should have clear meaning
- Clarity test:
  - Could a new person unambiguously specify value of all nodes and states?
  - “Dissolved oxygen is high” does not pass clarity test
  - “Dissolved Oxygen ≥ 90% saturation” passes clarity test

Adding States to Parent

- Original model: Default in Netica is 1 state
- Define states
  - Qualitative {low, medium, high, severe}
  - Quantitative {0 – 0.5, 0.5 – 2.0, 2.0 – 4.0, 4.0 – 10 ppm}
  - Or both {low = 0 – 0.5, medium = 0.5 – 2.0, high = 2.0 – 4.0, severe = 4.0 – 10 ppm}
How to Define States

1. Where possible, use existing / recognised / documented classifications – especially for physical parameters
2. Identify states using expert opinion & the needs of decision makers (decision thresholds)
3. Use percentiles of data / ranges within data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of input data</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissolved Oxygen</td>
<td>Expert knowledge (workshop II)</td>
<td>Extreme Low (0 – 40 %)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal (40 – 110%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extreme High (110 – 200%)</td>
</tr>
<tr>
<td>Salinity</td>
<td>Clunie et al. (2002) - fish data only</td>
<td>Low (0 – 1000 mg/L)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (1000 – 5600 mg/L)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (5600 – 10000 mg/L)</td>
</tr>
<tr>
<td>Turbidity</td>
<td>Expert knowledge (workshop II)</td>
<td>Low (0 – 100 NTU)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (100 – 1000 NTU)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (1000 – 10000 NTU)</td>
</tr>
<tr>
<td>Temperature modification</td>
<td>Ryan et al. (2001) Modelling of natural temperatures, and relating to temperatures required for spawning (Koehn and O'Connor 1990)</td>
<td>No Change (0 – 2°C)</td>
</tr>
<tr>
<td>(°C) from natural</td>
<td></td>
<td>Moderate (2 – 4°C)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Major (4 – 10°C)</td>
</tr>
</tbody>
</table>
Workshop Exercise

Defining states

Notes.................................
Module 4.
Conditional Probability Tables

*Expert elicitation*

---

**Estimating Probability Distributions**

- Probability distributions can be determined using information:
  - Elicited from experts (stakeholders)
  - Existing *process models*
  - *Mathematical* representations
  - *Statistical* relationships
- Uncertainties associated with each relationship is quantified in the probability distribution
Probability Distributions

- Categorical
  - E.g. Low, High

- Continuous Probability Distributions (CPD)
  - Dependent on software available and quality of information
  - Discretise nodes
    - assign states and specify distributions (expert elicitation and/or automated data learning)
    - Netica: Continuous Probability Tables (CPTs)
    - Low (1 - 5 mg/L), High (5 - 50 mg/L)

- True continuous distributions
  - Distribution type (eg. Log normal, exponential)
  - Moments of distribution: Mean, Standard deviation

Discretising a Continuous Probability Distribution

Histogram of C1
Lognormal

Loc 0.00750
Scale 1.026
n 100
Probability Distributions: Defining States

Netica models: First we need to specify states for each variable

Probability Distributions: Parameter Estimation

Parent variables: Marginal probabilities
Child variables: Conditional probabilities

Netica models: Next we need to assign some probabilities to our variables
Series of scenarios:

Given, retention time is short, light / phosphorous limitation is low, and temperature is low, what is the probability of the *Anabaena* algal population being characterised as Extremely Low, Low, Medium and High?

How confident are you in your prediction?

Probability Elicitation: Responses

Given, retention time is short, light / phosphorous limitation is low, and temperature is low, what is the probability of the algal population being characterised as Extremely Low, Low, Medium and High?

Expert Elicitation

Series of scenarios within CPTs:
1. Do the worst case scenario
2. Do the best case scenario
3. Are there any rules you can derive for parameters (e.g. if DO is < 30%, WQ = bad)
Problem: Expert Fatigue

• Problem with all elicitation methods is expert fatigue
• Advice
  – Use rules where possible (e.g. if DO is < 10% saturation, all fish dead)
  – Elicit the extremes of the distribution
  – Elicit medians
  – Use sensitivity analysis (later slides) to identify critical relationships in the model and if necessary, revisit elicitation

Problem: Unhappy Experts

• Experts can become dissatisfied with the elicitation process
• Advice:
  – Talk through this
  – Use subjective probabilities (models deterministic): keep model qualitative (Low, High)
  – Use large uncertainty bounds,
  – Or use flat priors (Low = 0.5, High = 0.5)
    • Reflect absence of knowledge
Documentation

• To document expert opinion, answer:
  – Who? When?
  – Limitations of expert knowledge
  – Limitations in model
  – Becomes: Knowledge gap, priority data need.

Feedback/Review Discussion

• Benefits of the approach?
• Any confusion/uncertainty to address tomorrow?
• Key concerns about the process and products?
Notes..........................................

Notes..........................................

© Integrated Catchment Assessment and Management (iCAM),
The Australian National University, 2007
Bayesian Network Training Workshop: Day 2

Landscape Logic

Overview of Day 1

- Introduction to the background of BNs
  - The process to build a BN
    - Conceptual models
    - Influence diagrams
    - Inputs into BNs (defining states and input data)
    - Evaluating BNs
    - BN software packages
What is the objective of the model

Conceptual model of how the system works

Transform conceptual model in influence diagram

Describe the model variables (states)

Populate & document assumptions and information used in BNs

Evaluation models (sensitivity and accuracy)

Testing model scenarios

Updating models

Developing our own Bayesian Networks

Day 1

Day 2

Workshop Exercise

Defining CPTs using expert elicitation
Module 4.
Conditional Probability Tables
continued

Quantitative data

Data Derived

Data sources: Monitoring, Research or Modelled

Text files (in Excel) that are ‘learnt’ by Netica:

Example file (Algal model)

Probability distributions assigned using frequency of observation in a file (using algorithms based on Bayes’ theorem)

Data can also be used to supplement expert elicited nodes (apply Bayes’ theorem)
### Conditional probability table for Algal network

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Phosphorus</th>
<th>Light</th>
<th>Retention Time</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>Short</td>
<td>88.5, 2.5</td>
<td>14.7, 2.7</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>Long</td>
<td>85.1, 2.5</td>
<td>5.7, 2.7</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>Short</td>
<td>24.6, 2.5</td>
<td>15.8, 2.7</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>Long</td>
<td>91.3, 2.5</td>
<td>7.2, 2.7</td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>Short</td>
<td>81.3, 2.5</td>
<td>21.4, 2.7</td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>Long</td>
<td>91.3, 2.5</td>
<td>7.2, 2.7</td>
</tr>
<tr>
<td>Medium</td>
<td>Low</td>
<td>Short</td>
<td>21.4, 2.5</td>
<td>23.1, 2.7</td>
</tr>
<tr>
<td>Medium</td>
<td>Low</td>
<td>Long</td>
<td>10.9, 2.5</td>
<td>83.2, 2.7</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Short</td>
<td>31.6, 2.5</td>
<td>68.3, 2.7</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Long</td>
<td>31.6, 2.5</td>
<td>68.3, 2.7</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>Short</td>
<td>21.4, 2.5</td>
<td>23.1, 2.7</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>Long</td>
<td>10.9, 2.5</td>
<td>83.2, 2.7</td>
</tr>
<tr>
<td>Extreme</td>
<td>High</td>
<td>Short</td>
<td>64.1, 2.5</td>
<td>35.9, 2.7</td>
</tr>
<tr>
<td>Extreme</td>
<td>High</td>
<td>Long</td>
<td>64.1, 2.5</td>
<td>35.9, 2.7</td>
</tr>
<tr>
<td>Extreme</td>
<td>Low</td>
<td>Short</td>
<td>7.6, 2.5</td>
<td>92.4, 2.7</td>
</tr>
<tr>
<td>Extreme</td>
<td>Low</td>
<td>Long</td>
<td>7.6, 2.5</td>
<td>92.4, 2.7</td>
</tr>
</tbody>
</table>

**Day Length**

- Short: 14.8 ± 0.6
- Long: 24.6 ± 0.8

**Temperature**

- Low: 5.46 ± 2.5
- Medium: 44.1 ± 6.8
- High: 43.7 ± 6.8
- Extreme: 6.71 ± 6.8

**Phosphorus/Light**

- High: 63.5 ± 2.5
- Medium: 11.3 ± 2.5
- Low: 25.3 ± 2.5

**Retention Time**

- Short: 48.0 ± 0.6
- Long: 52.0 ± 0.6

**Population**

- Ext. Low: 64.0 ± 2.5
- Low: 17.3 ± 2.5
- Medium: 19.2 ± 2.5
- High: 9.5 ± 2.5

**Phosphorus Limitation**

- Yes: 5.2 ± 0.3
- No: 94.8 ± 0.7

**Light Limitation**

- Yes: 0.8 ± 0.3
- Some: 18.8 ± 0.5
- No: 80.4 ± 0.7

**Flow**

- Ext. Low: 34.0 ± 0.6
- Low: 20.0 ± 0.6
- Medium: 16.6 ± 0.6
- High: 29.4 ± 0.6

Webb and Pollino
Deriving Parameters Using Data in Netica

- Lauritzen and Spiegelhalter (Bayes’ rule), EM algorithm (missing data)
- Less known about a variable, the greater the predictive uncertainty, which is reflected in the probability distribution
  - E.g. less frequent a parameter is measured

### Reporting: Parameterisation

<table>
<thead>
<tr>
<th>Subnetwork</th>
<th>Node Name</th>
<th>Node Type</th>
<th>Source Model</th>
<th>Data Source – Monitoring Report</th>
<th>Equation in Bayesian Network</th>
<th>Expert Input – Calibration Report</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crop/Livestock</td>
<td>Grain</td>
<td>Data</td>
<td>Grain Model</td>
<td>(2002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Livestock</td>
<td>Data</td>
<td>Livestock Model</td>
<td>(2003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fertiliser</td>
<td>Data</td>
<td>Fertiliser Model</td>
<td>(2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Irrigation</td>
<td>Data</td>
<td>Irrigation Model</td>
<td>(2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water Quality</td>
<td>Data</td>
<td>Water Quality Model</td>
<td>(2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water Use</td>
<td>Data</td>
<td>Water Use Model</td>
<td>(2007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wetlands</td>
<td>Data</td>
<td>Wetlands Model</td>
<td>(2008)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wildlife</td>
<td>Data</td>
<td>Wildlife Model</td>
<td>(2009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wetland Ecosystem</td>
<td>Data</td>
<td>Wetland Ecosystem Model</td>
<td>(2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Integrated Catchment Assessment and Management (ICAM), The Australian National University, 2007
Tip:
A model is only as reliable as its least reliable input

Jørgensen and Bendoricchio

Workshop Exercise

Identify potential data sources for your BN
Notes

Notes

Notes

Notes
Module 5.
Model Evaluation Tools

Sensitivity analysis
Knowledge gaps
Predictive accuracy

Model evaluation ⇒ Model uncertainties / Ranking risks

Case Study example:
Native fish in the Goulburn Catchment
- Data and expert-driven
Fish Network

- 5 sub-networks
- Water Quality
- Flow
- Structural Habitat
- Biological Interactions

- 2 query nodes
- Fish Abundance
- Fish Diversity

- 23 sites
- 6 reaches
- 2 temporal scales
- 1 and 5 year changes

Pollino, Woodberry, Nicholson, Korb, Hart

Model Evaluation Tools

- Quantitative
  - Sensitivity Analyses*
  - Predictive Accuracy
- Qualitative
  - Expert
  - Real data vs. Model Prediction

* Metrics: E.g. Entropy, Mutual Information, Variability across states
Model Evaluation

• Elicitation review
  – Review variable and state definitions
    • Clarity test, consistency
  – Review graph structure
  – Review probabilities
    • Compare probabilities with each other

• Sensitivity analysis
  – Measures effect of one variable on another
  – Compare with expert intuition to evaluate model
  – Evaluate whether additional modeling is needed

• Case-based evaluation (Predictive accuracy)
  – Run model on set of test cases
  – Compare with expert judgment or “ground truth”

Quantitative Evaluation

• Sensitivity Analysis
  ▪ Assist in identifying variables and causal relationships of importance
  ▪ Is there problems with your model structure or quantitative relationships?
  ▪ Where is more effort needed to better quantify relationships?
    • Consult experts
    • Acquire more data (research / monitoring)
    • Identify key knowledge gaps (priority risk?)
Outputs from Sensitivity Analysis

1. Prioritise key variables
2. Identify sensitive variables that need further quantification (i.e. knowledge gaps / priorities for data collection)
3. Evaluation: Does the model make sense?
4. Simplifying models – which nodes are unimportant

Prioritisation of Risks

• Sensitive Variables
  ▪ If a sensitive variable is accurately quantified, it represents a PRIORITY RISK
  ▪ If sensitive variable cannot be more accurately quantified based on the information available, it represents a KEY KNOWLEDGE GAP and is considered a PRIORITY RISK
Model Robustness Tools

- Sensitivity Analysis
  - Is there problems with your model structure?
  - Is there errors in the quantitative relationships?
  - Assist in identifying variables of importance
  - Where is more effort needed to better quantify relationships?
    - Consult experts
    - Acquire more data
    - Identify key knowledge gaps and priority risks

Sensitivity Analysis

- Sensitive influences in a BN can be quantified using two types of measures: entropy and mutual information.
- These can be used to rank a specified set of interest nodes.
- Model review: is a variable either too sensitive or insensitive to other variables in particular contexts? This may help identify errors in either BN structure or the CPTs.
How to do Sensitivity Analyses

• Metrics:
  – Entropy
  – Mutual Information

• Displaying results graphically

Sensitivity Analysis

• Entropy, $H$, is commonly used to evaluate the uncertainty or randomness of a variable ($X$) characterised by a probability distribution, $P(x)$:

$$H(X) = - \sum_{x \in X} P(x) \log P(x)$$

• Entropy measures assess the average information required in addition to the current knowledge to specify a particular alternative (Das 2000).
Sensitivity Analysis

Entropy of Future Abundance: 0.386506
Future Diversity: 0.046391
Water Quality Habitat Descriptor: 0.01806
Hydraulic Habitat Descriptor: 0.00818
Natives Biological Potential Descriptor: 0.00073
Temperature Modification: 0.00044
Barrier: 0.00043
Change in Avr Flows Summer-Autumn: 0.000231

Sensitivity Analysis

• Mutual information (I) is used to measure the effect of one variable (X) on another (Y):

$$I(X,Y) = H(X) - H(X|Y)$$

• where I(X,Y), is the mutual information between variables X and Y.
• This measure reports the expected degree to which the joint probability of X and Y diverges from what it would be if X were independent of Y (If I(X,Y) is equal to zero, X and Y are mutually independent).
Sensitivity Analysis

<table>
<thead>
<tr>
<th>Node</th>
<th>MI</th>
<th>Rank</th>
<th>MI</th>
<th>Rank</th>
<th>MI</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>FutureDiversity</td>
<td>0.001150</td>
<td>1</td>
<td>0.000955</td>
<td>1</td>
<td>0.001650</td>
<td>1</td>
</tr>
<tr>
<td>WQmain</td>
<td>0.007922</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Site</td>
<td>0.029622</td>
<td>3</td>
<td>0.000767</td>
<td>4</td>
<td>0.005369</td>
<td>5</td>
</tr>
<tr>
<td>OverallFlow</td>
<td>0.028998</td>
<td>4</td>
<td>0.002151</td>
<td>2</td>
<td>0.025601</td>
<td>3</td>
</tr>
<tr>
<td>BiolPoten</td>
<td>0.025654</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>0.040023</td>
<td>2</td>
</tr>
<tr>
<td>Barrier</td>
<td>0.023402</td>
<td>9</td>
<td>0.000400</td>
<td>8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Temp</td>
<td>0.023402</td>
<td>7</td>
<td>0.000294</td>
<td>10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AvrSummer</td>
<td>0.022223</td>
<td>8</td>
<td>0.000186</td>
<td>7</td>
<td>0.005937</td>
<td>9</td>
</tr>
<tr>
<td>Type</td>
<td>0.022221</td>
<td>9</td>
<td>0.000389</td>
<td>9</td>
<td>0.002369</td>
<td>10</td>
</tr>
<tr>
<td>OverallSummer</td>
<td>0.020855</td>
<td>10</td>
<td>0.000642</td>
<td>5</td>
<td>0.008152</td>
<td>6</td>
</tr>
<tr>
<td>OverallWinter</td>
<td>-</td>
<td>-</td>
<td>0.000865</td>
<td>3</td>
<td>0.004622</td>
<td>2</td>
</tr>
<tr>
<td>AvrWinter</td>
<td>-</td>
<td>-</td>
<td>0.000195</td>
<td>6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PopStat</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.008181</td>
<td>7</td>
</tr>
<tr>
<td>MinSummer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.005469</td>
<td>5</td>
</tr>
</tbody>
</table>

Sensitivity Analysis

- Outputs can also be represented graphically showing the change in the probability of a variable when parent variables are altered over the parameter space (from a probability of 0 to 1)
- Graphical representations are often more widely understood by audiences
Predictive Accuracy

- Quantitative analysis
  - Data split (80% training - 20% testing)
    - Future abundance error rate = 10.71%
    - Future diversity error rate = 0% (all low)
  - Withhold case data from training, and use this to predict endpoints – errors calculated on percentage of error rates incorrectly predicted
Confusion matrix

<table>
<thead>
<tr>
<th>Actual</th>
<th>0 to 12</th>
<th>12 to 44</th>
<th>44 to 101</th>
<th>101 to 155</th>
<th>155 to 311</th>
<th>311 to 813</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 12</td>
<td>1797</td>
<td>127</td>
<td>28</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12 to 44</td>
<td>297</td>
<td>465</td>
<td>88</td>
<td>10</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>44 to 101</td>
<td>64</td>
<td>162</td>
<td>189</td>
<td>23</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>101 to 155</td>
<td>8</td>
<td>35</td>
<td>63</td>
<td>38</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>155 to 311</td>
<td>2</td>
<td>13</td>
<td>40</td>
<td>15</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>311 to 813</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>22</td>
<td>4</td>
</tr>
</tbody>
</table>

Qualitative Evaluation

- Qualitative
  - Expert evaluation: How realistic is the model
  - Historic data vs. Model Prediction

Relative abundance

P(Native Fish) = High
Tip:

An important outcome of your model may be a better understanding of your system, and not a reliable, quantitative prediction.

Jørgensen and Bendoricchio

Summary Steps for Models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert opinion only</td>
<td>Expert elicitation</td>
<td>Sensitivity analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expert review</td>
</tr>
<tr>
<td>LOTS of data</td>
<td>Data learning (L&amp;S or EM algorithm)</td>
<td>Sensitivity analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Predictive accuracy</td>
</tr>
<tr>
<td>Expert opinion and some data</td>
<td>Expert elicitation</td>
<td>Sensitivity analysis</td>
</tr>
<tr>
<td></td>
<td>Data learning (L&amp;S or EM algorithm)</td>
<td>Predictive accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expert review</td>
</tr>
</tbody>
</table>
Limitations of BNs

What do they do Poorly?

- Dynamic relationships
  - Supported by some packages
- Large-scale networks
  - What is the right graph?
  - Often more effort is put into CPDs c.f. structure – if structure is flawed, model is flawed
- Not all packages support true CPDs
- Limited interactions
  - Specifying CPTs with ‘experts
What do they do Poorly?

- Problems associated with expert elicitation
  - Intensive process (esp. when model is complex)
  - Consensus vs. multiple models?
  - How many experts is enough?
- Potential to be abused - user-friendly nature
- Used to expedite the responsibility of the decision maker

Workshop Exercise:

*Finalise draft BNs*
Final Discussion

Feedback

Where to from here?

Workshop evaluation

Feedback/Review Discussion

• Benefits of the approach?
• Any confusion/uncertainty to address in the future?
• Key concerns about the process and products?
Where to from Here?

• Take your learnings from today back to your region group.
• Build and/or contemplate BNs as it is suitable within your group.
• iCAM staff will visit each region individually (1 day each) in May and again in August.
• In between times please contact iCAM staff with questions, queries or problems.

• Please complete an evaluation form provided.

• Thanks
For Technical Support at iCAM

- Carmel Pollino:
  - Carmel.Pollino@anu.edu.au
  - (02) 6125 8132
- Jenifer Ticehurst:
  - Jenifer.Ticehurst@anu.edu.au
  - (02) 6125 6751
- Matthew Tighe:
  - Matthew.Tighe@anu.edu.au
  - (02) 6125 9019