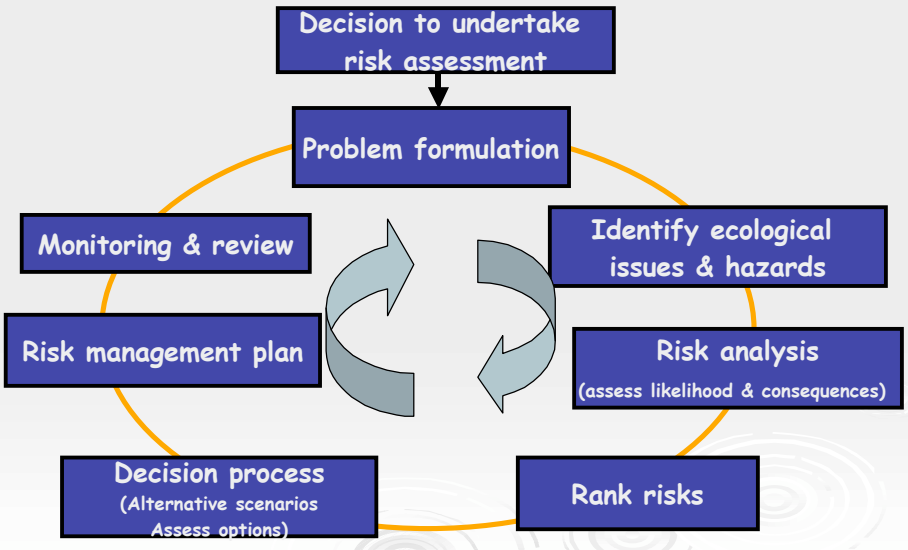
Quantifying Ecological Risks using Bayesian Networks:
Modelling in an Uncertain World

Carmel A. Pollino

Water Studies Centre

- Bayesian networks and the ERA process?
- What is a Bayesian network?
- Steps for building a Bayesian network using case study examples
- Workshop exercise

Risk Management Process



Hart et al. (2005)

Bayesian Networks

- belief network
- causal network
- probabilistic network
- knowledge map

Rev. Thomas Bayes
(1702-1761)

BNs and ERA

- Enable integrated decision-making
  - Investigates multiple factors / hazards affecting endpoint(s)
  - Draw together disparate data sets / existing models
- Promote:
  - **Iterative** and **Adaptive** approaches to environmental management
  - Transparent and tractable decision-making
  - Inform future monitoring and targeted research

BNs and ERA

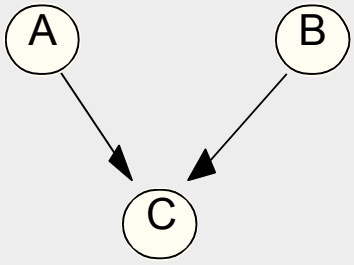
- Facilitate stakeholder engagement (technical and non-technical)
  - Communication
  - Adoption
- Complexity
  - Multiple stressor
- Address issues of uncertainty

Sources of uncertainty

- **Statistical variation** (eg. parameter measurements)
- **Subjective judgments** (of model structure and estimation of parameters)
- **Inherent variability** (as model is general and as knowledge is incomplete)
- **Inherent randomness** (represent unknown variables of causal relationships)
- **Disagreement**

What is a Bayesian Network?

- Two components
  - Structure
  - Probabilities
- Links between variables represent causal relationships (as with a conceptual model)



- Probabilistic relationships are used to describe the strengths between variables
- Analysis of uncertainty
- Inputs: expert opinion (technical / non-technical), literature, **monitoring / research data, other models**

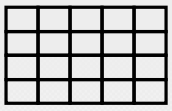
Definitions

- **Prior probability**
  - Likelihood the parameter will be in a particular state
- **Conditional probability**
  - Calculates the likelihood of a state, given the states of the input parameters
- **Posterior probability**
  - The likelihood that a parameter will be in a particular state given the input parameters, the conditional probabilities, and rules governing how probabilities combine

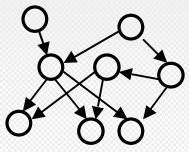
Components of Bayesian Network

1. set of random variables (nodes)
   discrete, categorical, and continuous variables

2. set of directed links/arrows connect the nodes to represent dependencies
   Note: absence of arrows indicate independence



3. a conditional probability table is associated with each node (prior probability)



4. Graph with no directed cycles
   (DAG: directed acyclic graph - graph theory)

Updating using Bayes' Rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

where P(A) is the prior distribution of parameter, A. After collection of data B, P(A|B) represents the posterior distribution, given the new knowledge

How do I build a BN?

- Software platform:
  - Today we are using Netica
- Define model scope
  - Spatial and temporal
- Specify structure
  - Expert Elicited
  - Automated (Structure learning)
- Specify probability distributions
  - Priors (flat, expert, other models or data)
  - Inference algorithms (L&S, EM, GD)

How do I build a BN?

- Test model robustness
  - Predictive accuracy
  - Sensitivity analysis
  - Expert evaluations
- Revise model and Test
  - Structure and CPDs
- Use for inference and prediction
- Update structure and CPDs

Model scope

- Answer management / policy relevant questions directly
  - Eg. What conditions are required in the Goulburn main channel to achieve sustainable native fish populations?
  - Spatial and temporal scales

Model Development

The Real World

Conceptual Model

Bayesian Network

Building Bayesian Networks

Model made up of 2 parts:

1. **Graphical component (Structure)**
2. **Quantitative component (Probabilities)**

Test model accuracy

Test scenarios

Conceptual models ⇒ Quantitative Models

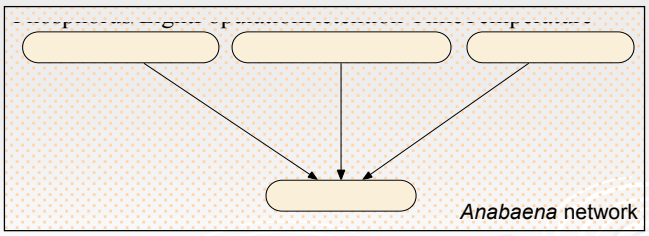
Case Study example:

Blue Green algal booms (*Anabaena*) in Burke Weir

- Data-driven model

Structural Development

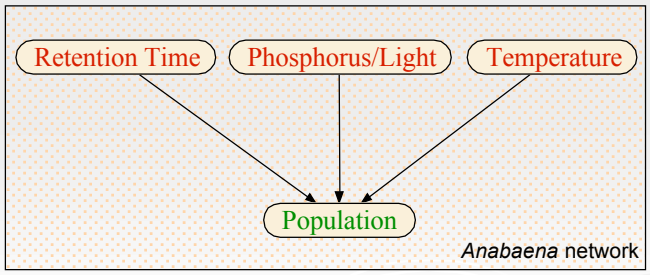
- Automated Learning from data (MML)
- Expert Elicitation of conceptual model
- Example: Diverse habitat structure for fish in a riverine environment



Structural Development

Model viewed as a graph

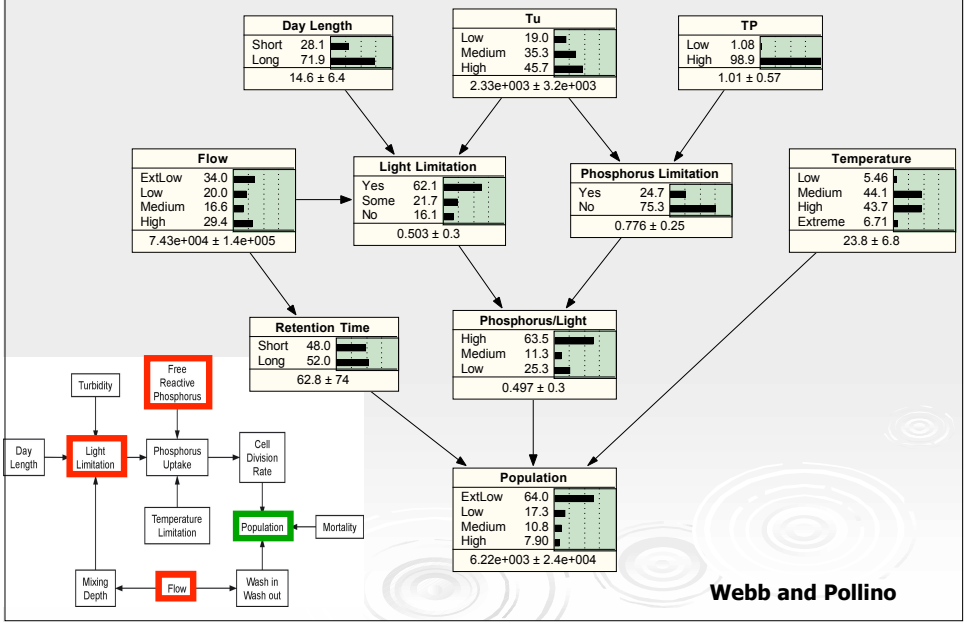
- Round nodes are important system variables
- Causal dependencies represented by arrows



Independent variables: marginal probability distributions

Dependent variable: outcome dependent on parent variables, integrated to form a conditional probability distribution

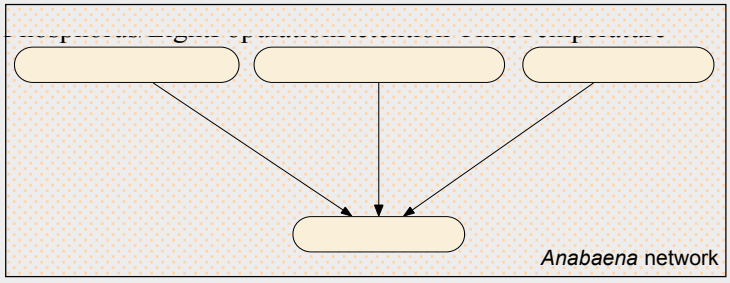
Anabaena Network



Probability Distributions

- Discrete (categorical)
- Continuous Probability Distributions (CPD)
  - Dependent on software available and quality of information
  - Discretised CPD into ranges (Knowledge engineer, automated)
    - Netica: Continuous Probability Tables (CPTs)
  - True continuous distributions

Bayesian Networks



Graphical model implies:

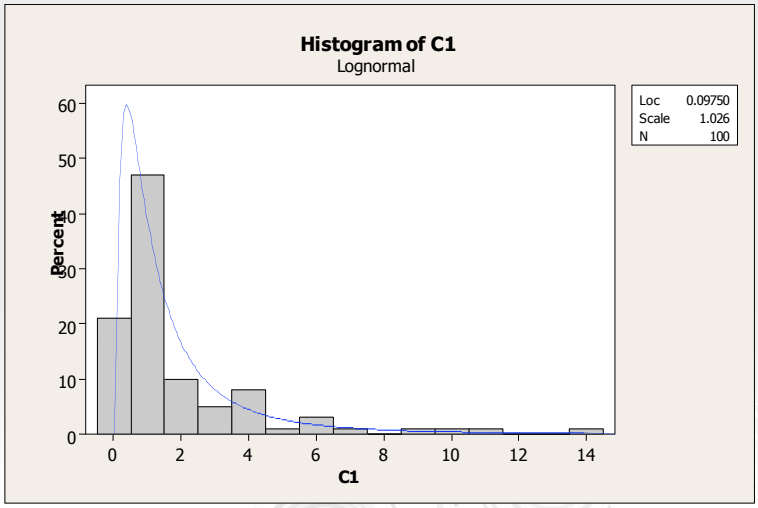
P (RT,PL,T,Pop) = P(Pop|RT,PL,T)P(RT)P(PL)P(T)

- Joint distribution of variables

Specifying Conditional Probabilities

- Conditional probability distributions can be determined using:
  - Elicited from experts
  - Existing process models
  - Mathematical representations
  - Statistical relationships
- Uncertainties associated with each relationship is quantified in the probability distribution

Continuous Probability Distribution



Expert Elicitation*

Series of scenarios:
Given that water quality is high, change in flow regime is low, structural habitat quality is high, and biological potential is low, what is the probability of fish abundance being characterized as "Low"? "High"?

How confident are you in your prediction?

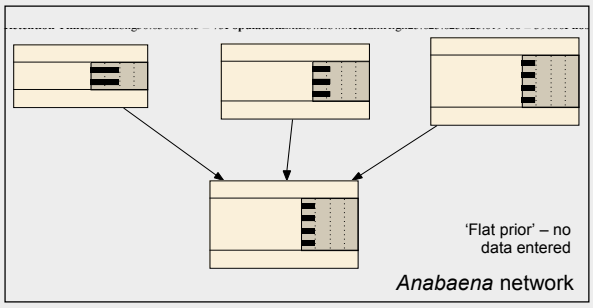
Take into consideration changes to the outcome nodes given the temporal scales of 1 year and 5 years.

* See Morgan, M. G. and M. Henrion (1990). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge, UK, Cambridge University Press.

Parameter estimation (Netica)

- Expert and/or data derived
- Lauritzen and Spiegalhalter (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

Assign Probabilities



- Use Continuous distributions (eg. Log normal, exponential, etc.) OR Discretise nodes – assign states and specify distributions (expert elicitation and/or automated data learning)

Netica - [Pop Table (in net Algal_model_1___equal_cpts_EM_)]

File Edit Table Window Help

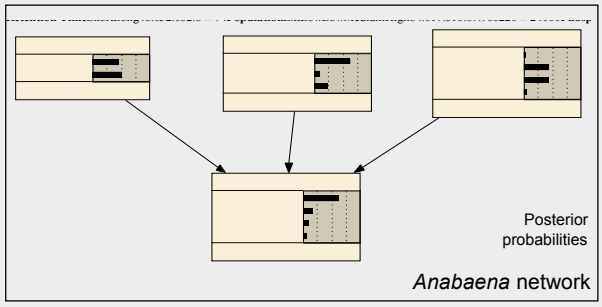
Node: Pop Apply Okay

Chance Percentages Reset Close

Temperature	Phosphorus/Light	Retention Time	ExtLow	Low	Medium	High
Low	High	Short	80.522	14.074	2.775	2.629
Low	High	Long	88.325	3.763	3.961	3.951
Low	Medium	Short	24.989	25.063	25.175	24.773
Low	Medium	Long	85.295	4.984	4.843	4.878
Low	Low	Short	23.474	23.628	29.810	23.088
Low	Low	Long	31.383	22.662	20.699	25.256
Medium	High	Short	89.128	5.606	4.851	0.415
Medium	High	Long	86.522	8.691	3.799	0.988
Medium	Medium	Short	27.809	27.157	26.889	18.144
Medium	Medium	Long	60.383	1.400	1.914	36.302
Medium	Low	Short	8.930	7.787	77.266	6.017
Medium	Low	Long	60.257	18.574	17.093	4.076
High	High	Short	65.304	23.146	9.815	1.735
High	High	Long	68.533	27.034	1.560	2.873
High	Medium	Short	21.228	24.561	14.108	40.103
High	Medium	Long	15.766	42.016	21.618	20.601
High	Low	Short	22.281	24.301	23.015	30.403
High	Low	Long	53.324	19.757	10.398	16.521
Extreme	High	Short	42.818	44.086	11.124	1.972
Extreme	High	Long	78.089	5.284	5.783	10.844
Extreme	Medium	Short	25.620	21.304	32.030	21.046
Extreme	Medium	Long	25.828	16.166	14.482	43.524
Extreme	Low	Short	9.433	8.291	74.067	8.209
Extreme	Low	Long	6.685	4.608	40.512	48.195

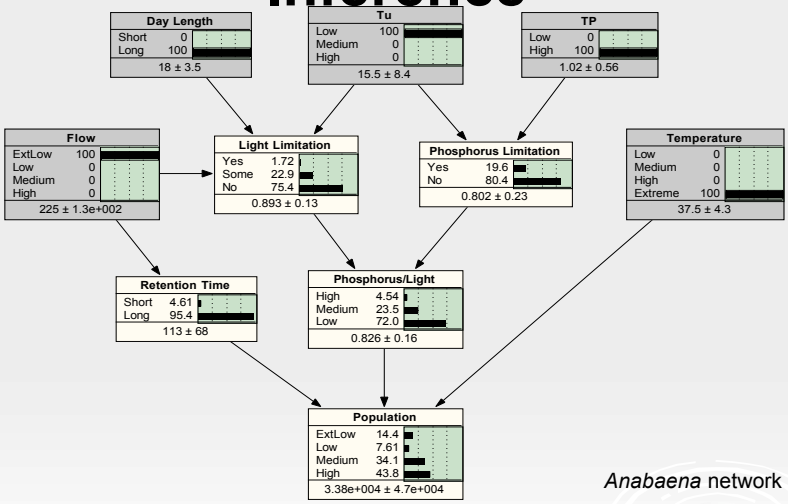
Anabaena network

Solved Network

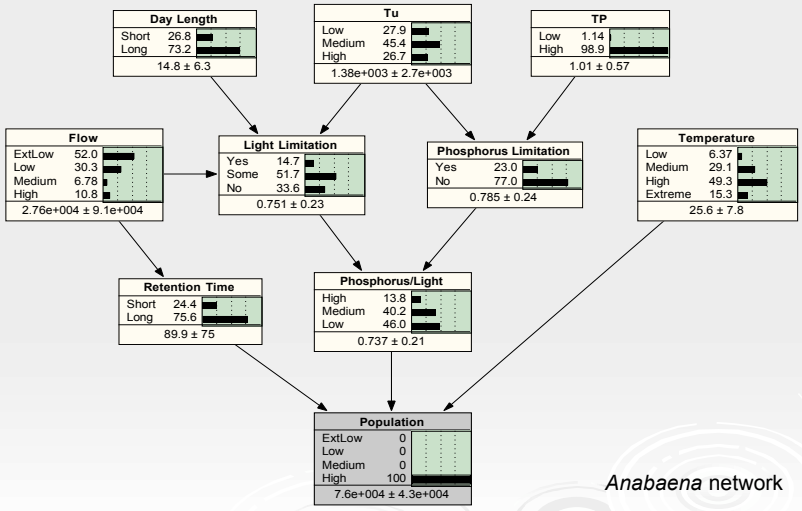


Given the input parameters, the conditional probabilities, and any rules governing how the probabilities combine - the network is solved.

Inference



Diagnostic



Anabaena network

Model evaluation ⇒ Model uncertainties / Ranking risks

Case Study example:

Native fish in the Goulburn Catchment

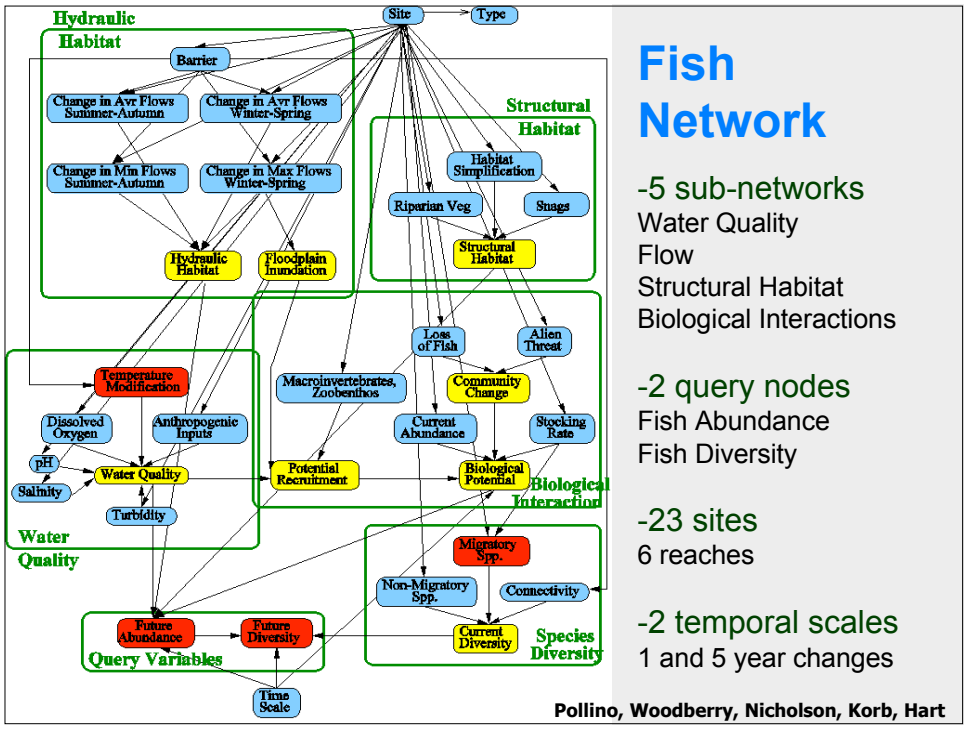
- Data and expert-driven

Tools Available

- 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?)

Tools Available

- Predictive Accuracy tests
  - Real vs. Model data
    - Data splits
- Plot predicted output vs. historical data
- Sensitivity analysis
- Expert evaluation
  - Is the model outputs reasonable?



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

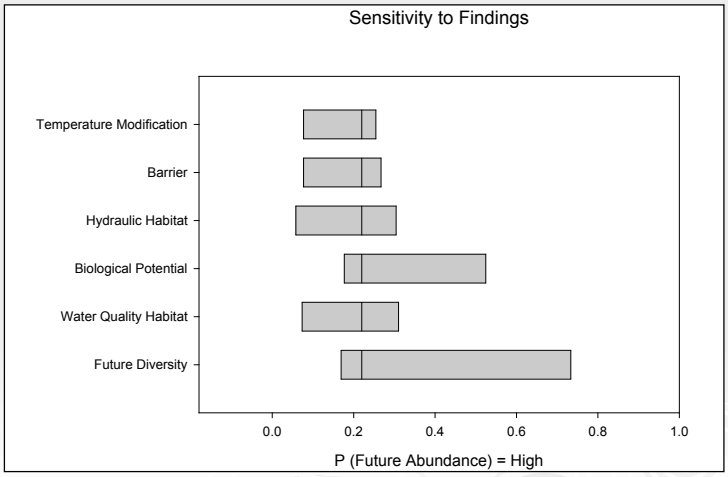
Model Evaluation tools

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

Quantitative Evaluation

- Predictive Accuracy
  - Data split (80% training, 20% testing)
  - Error Rate (Future Abundance) = 5.8%
- Limited data
  - Lack of variability in abundance of fish communities throughout catchment (mostly low – poor condition) ⇒ Expert evaluation

Ranking Risks - sensitivity



Qualitative Evaluation

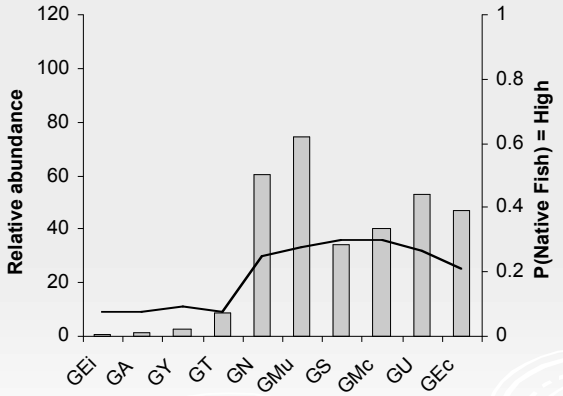


Figure 5: Relative Abundance Data (left axis - bars) versus BN Model Predictions (right axis - line) for Sites in the Goulburn Main Channel.

Complex Models ⇒ Simple Models

Case Study example:

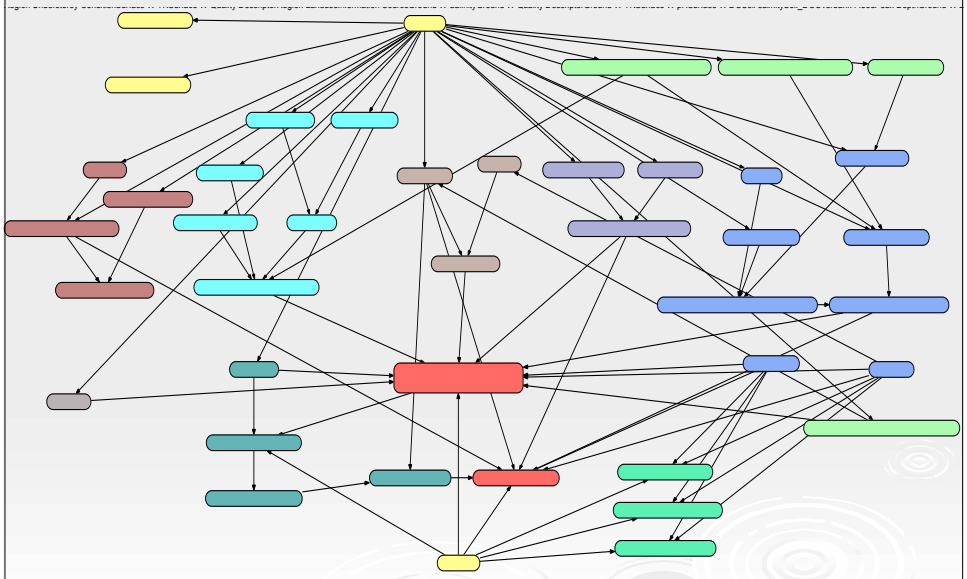
Black box *(Eucalyptus largiflorens)* depressions on the Murray floodplain

- Data and expert-driven

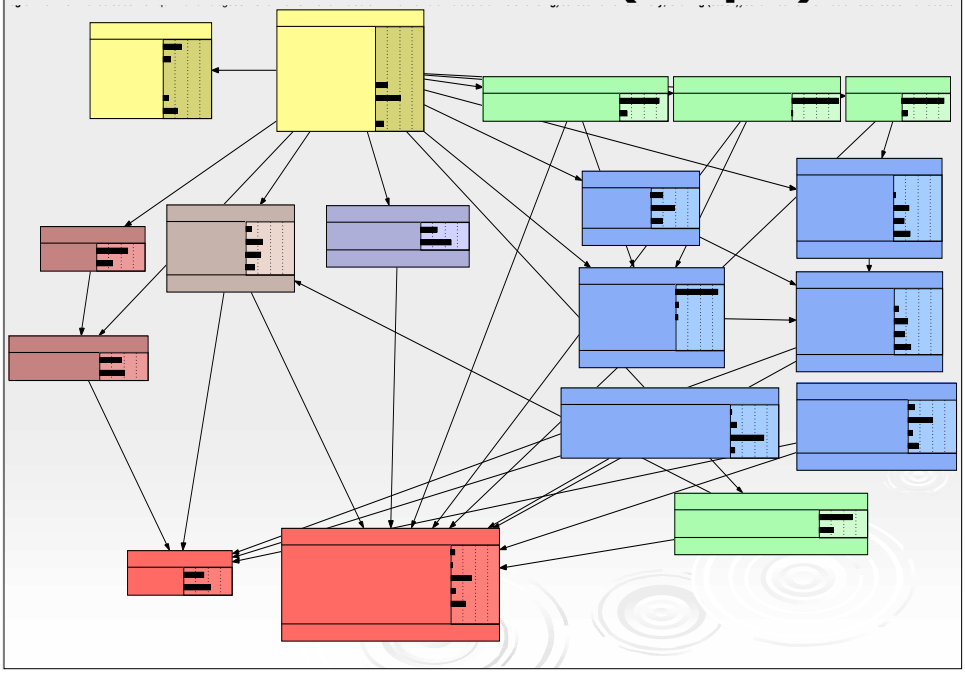
"A theory should be as simple as possible, but no simpler."

Albert Einstein

Black box Network (complex)

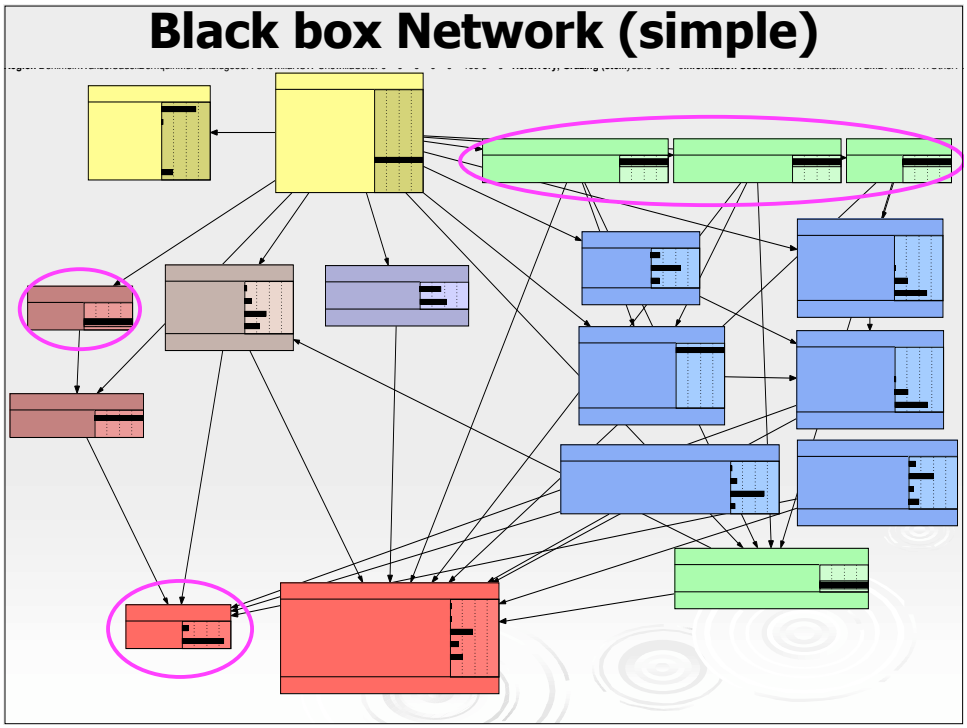


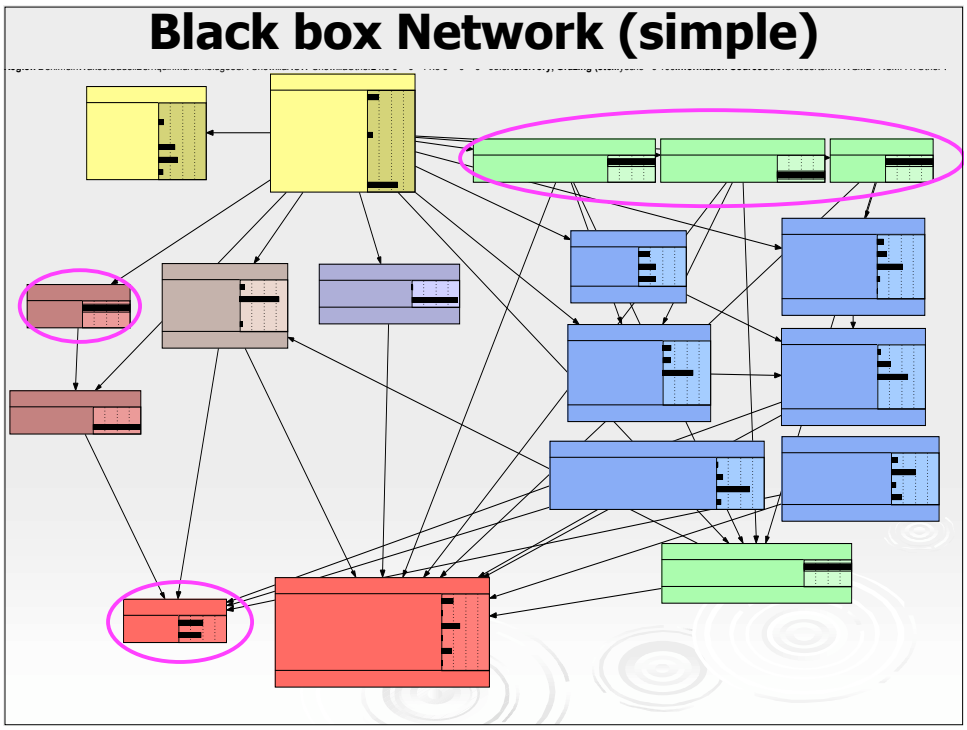
Black box Network (simple)



Testing Scenarios

- Management or system changes
  - Examine relative change of probabilities which conveys expected system response while accounting for predictive uncertainties
  - Predictions: Historic, now, future


Black box Network (simple)


Black box Network (simple)

✓**Bayesian Networks** are more than just qualitative tools ⇒ **powerful quantitative tools**

✓Fit into a cycle of **adaptive management / continual learning**

✓Range of flexible **evaluation** tools

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

BN Packages

- Analytica
- BayesiaLab
- Bayes Net Toolbox
- Deal (package available in R)
- Genie
- Netica
- WinBUGs

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

Workshop exercise (Netica)