Developing Bayesian network models within a Risk Assessment framework

C. A. Pollino

Barry T. Hart

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Developing Bayesian network models within a Risk Assessment framework

Carmel A. Pollino[1] and Barry T. Hart[2]
[1]The Fenner School of Environment and Society, The Australian National University, Canberra, Australia, [2]Water Studies Centre, Monash University, Melbourne, Australia

Abstract: The risk assessment framework is increasingly being applied to examine both human and non-human stressors on ecological systems. Risk-based decision-making aims to quantify the likelihood of a threat occurring, the consequences of this to an ecological system, process or value, and the associated uncertainty in the predictions. Until recently, the ability to predict changes in dynamic ecosystems due to stressors was limited by both the poor understanding of the drivers of ecological processes and structure, and the lack of modelling tools that could represent such complexity with associated uncertainties. However, the recent growth in the use of Bayesian network tools for ecological risk assessments has resulted in major advances in better understanding and managing ecosystems despite their inherent complexity. Bayesian networks have the advantage of being able to investigate the impacts of multiple stressors in complex environments, while explicitly acknowledging the associated uncertainties resulting from inherent variability and lack of knowledge of ecological systems within an adaptive framework. Bayesian networks have the flexibility to incorporate diverse knowledge systems, ranging from 'gut feel' to quantitative process-based or simulation models. In this paper, we discuss the relationships between the risk assessment framework and Bayesian network building process, and will illustrate the main concepts with a series of Bayesian network models.

***Keywords**:* Bayesian network; risk assessment; ecology

1. INTRODUCTION

The development of quantitative models to support decision-making in environmental management is considered to be of high priority. Recently, ecological risk assessment and risk management approaches have arisen as an approach to improve integrated decision-making processes. The ecological risk assessment framework offers a formal approach to decision-making, allowing for a greater understanding of decision processes and how they relate to ecological endpoints to be achieved. Bayesian networks are being utilized to produce complex quantitative models that are both pragmatic and scientifically credible. In this paper, we overview the risk assessment and Bayesian networks cycles, and discusses the relationship between these cycles using two case studies.

1.2 Framework for undertaking a Risk Assessment

Risk is defined as a state of uncertainty where some of the possibilities involve a loss, catastrophe, or other undesirable outcome [Hubbard, 2007]. Consequently, uncertainty is a key component in assessing risk. In an ecological context, the most common sources of uncertainties include poor or incomplete understanding in the linkages between ecological processes within a system, and the inherent variability associated with such processes. As risk is the product of likelihood/hazard and consequences/effects, it is not a directly measurable attribute.

Risk assessment is the initial step in risk management. It involves the determination of risk related to a past or current situation and a recognised threat. It also provides a robust process that incorporates a transparent, scientific, precautionary and sustainable approach to the management of environmental risks.

The ecological risk assessment framework, involves a number of key steps:

- *Defining the problem* –Defining assessment objectives and scope, identifying and engaging stakeholder groups.
- *Defining important ecological assets, and identifying hazards to these assets* - hazards are identified in the context of ecological assets, activities that mediate hazards are also identified. A conceptual model/influence diagram is constructed, preferably in collaboration with stakeholder groups.
- *Analysing the risks to the ecological values* – the risk analysis process used needs to be appropriate for the situation in order to provide adequate information for decision-making. Both qualitative and quantitative information types should be incorporated in the risk analysis step. The assumptions, strengths and limitations of the analysis should be documented.
- *Characterising the risks* - the technical details of risk analyses needs to be made accessible to decision-makers and broader stakeholders. In particular, the uncertainties and assumptions associated with analyses require careful and transparent documentation.
- *Making decisions* – selection of the best management option or strategy will be the one that results in the effective minimisation of the ecological risks, while also being cost-effective and acceptable to the stakeholders. Guidance is provided on a number of multi-criteria methods for assisting this process.
- *Managing the risks* – a risk management plan provides recommendations on managing or mitigating all high or unacceptable risks. The risk management plan should include a robust program to monitor progress to ensure the strategies are working, and a review and feedback process for making changes if needed.

These steps are shown in Figure 1.

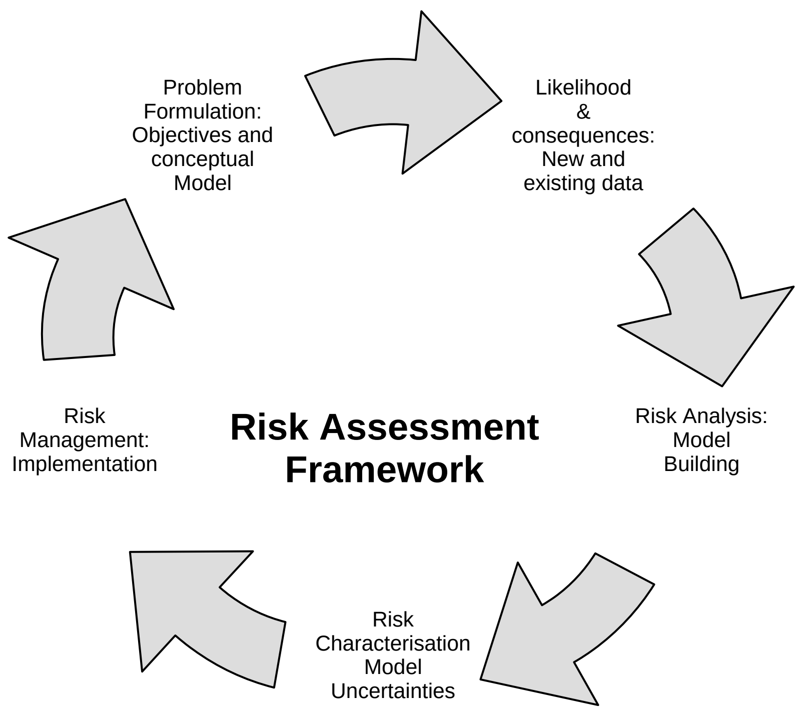

Figure 1: Risk Assessment Framework

1.1 Framework for constructing Bayesian networks

Bayesian networks (BNs), also known as Bayesian Belief Networks (BBNs) and Belief Networks, are increasing being used in ecological applications as they offer a pragmatic and scientific approach to modelling complex ecological systems where high uncertainties exist. Unlike many other ecological modelling approaches, BNs can utilise prior knowledge and data to model systems. Furthermore, BN models are particularly useful for analysing and communicating causal assumptions not easily expressed using mathematical notation, and for analysing multivariate and complex relationships among variables. Of particular note, as BNs promote stakeholder involvement, they fit into the participatory decision-making approach advocated in environmental decision-making. They also provide a platform in which disciplines can work together in a more integrative fashion.

BNs are graphical models that use probabilistic expressions to describe the relationships among variables. They are able to explore and display causal relationships between key factors and final outcomes of a system in a straightforward and easily understood manner. A *prior probability* represents the likelihood that an input parameter will be in a particular state. The *conditional probability* calculates the likelihood of the state of a variable given the states of input variables affecting it. And the *posterior probability* is the likelihood that a variable will be in a particular state, given the input variables, the conditional probabilities, and the rules governing how the probabilities combine. BNs use Bayes' Theorem to update or revise the beliefs of the probabilities of system states taking certain values, in light of new evidence.

As BNs are causal, they can also be used to calculate the effectiveness of interventions, such as management decisions, and system changes, such as those predicted for climate change. Importantly, the uncertainties associated with these causal relationships can also be explored at the same time. BNs are able to maintain clarity by making causal assumptions explicit [Stow & Borsuk, 2003] and are often used for modelling when relationships to be described are not easily expressed using mathematical notation [Pearl, 2000]. BNs use the network structure to calculate the probability certain events will occur, and how these probabilities will change given subsequent observations or a set of external (management) interventions.

1. Structure of a Bayesian network

The first step in constructing a BN is to develop a causal structure with relevant variables (nodes) and dependencies, which forms the model framework. Criteria for inclusion of variables in a BN are that the variable be: (a) manageable, (b) predictable, or (c) observable at the scale of the management problem [Borsuk *et al.*, 2004]. Any processes or factors not included become part of the uncertainty of the network, forming the predictive uncertainty described in probability distributions. Within a Bayesian network, sub-networks can be nested to describe physical or chemical processes relevant to the spatial scale specified.

2. Discretization of nodes (assigning states)

States or condition of the variables can be categorical, continuous or discrete. In order to represent continuous relationships in a Bayesian network, a continuous variable must be divided or discretized into states. The states of a variable can be numerical ranges (< 3, >3) or expressions (that can also represent data if appropriate, e.g. acceptable ≤ 3, unacceptable >3). If relevant, these states can represent targets, guidelines, existing classifications or percentiles of data.

3. Specification of prior probabilities

After defining states, linkages between nodes need to be described. Parent nodes lead into child nodes, the outcome of child nodes are conditional on how the parent variables combine. This is relationship is defined using conditional probability tables (CPTs). CPTs can be derived via one or a combination of methods:

- Direct elicitation of scenarios from expert;
- Parameterisation from datasets;
- Equations that describe relationships between variables.

4. Calculating posterior probabilities

Data or new knowledge can be incorporated into BNs and used to calculate posterior probabilities. Data sources can be entered into the network as a series of 'cases'. Cases can represent data collected during a monitoring exercise, undertaken as part of a research study, and so on.

5. Model evaluation

A range of validation tools can be used for BNs. Evaluation can involve data or technical experts, or both. Quantitative evaluation with data is preferable. Such measures include predictive accuracy and sensitivity analyses.

Predictive accuracy tests are used to determine model error rates, which are quantified using data (although not the same data used for model parameterisation). This method measures the frequency with which the predicted node state (that with the highest probability) is observed, relative to the actual value. Outcomes can be used to identify weaknesses in the model, and where more effort can be targeted in order to improve model accuracy.

Sensitivity analysis is used to identify the key drivers in the model and major knowledge gaps in our understanding. Sensitivity analysis of mathematical models can be used to investigate the uncertainties and inaccuracies in model structure, relationships and outputs, and subsequently identify where priority knowledge and data gaps exist. Thus, based on these results, recommendations for targeted monitoring and research studies can be made.
Sensitivity analyses provide a ranking of importance of variables, relative to the variable of interest (usually the endpoint). These variables indicate where better quantification in the network should be investigated and identify the most influential variables on model endpoints. Subsequently, these are the variables that should be given greater attention. In a management context, it is these variables that may represent key management actions or knowledge gaps. As sensitivity findings can differ for different spatial areas of interest or scenarios tested, key knowledge gaps and priority risks can also differ.

6. Knowledge gaps and priority risks

Having established the structure of the model, and the relationships used to drive the model, the key knowledge gaps in our understanding and priority risks can be identified. Sensitivity analysis can be used to examine where key uncertainties in model linkages exist.

7. Testing management scenarios

Management scenarios can be tested by entering new information into the network as evidence, directly changing the distribution of probabilities on the node itself.

2. BAYESIAN NETWORKS IN A RISK ASSESSMENT FRAMEWORK

Ecological risk assessment frameworks are increasingly being used as they add rigor to traditional natural resource management decision-making [Burgman, 2005; EPA, 1998] addressing limitations of traditional approaches to natural resource management, such as decisions being based on expert judgements alone, or on single focus assessments. Expert opinion (and weight of evidence approaches) are subject to cognitive and knowledge-based bias [Newman & Evans, 2002; Pollino & Hart, 2006]. Whereas assessments that only have a single focus or hazard relationships, also introduce bias by not integrating other system processes. Although such assessments are limited by the information available, they still undergo little review or updating.

Many natural resource models for decision-making still only poorly represent uncertainties associated with predictive outputs. Without characterisation of uncertainties and focus on ecological systems as a whole, the prioritisation of hazards and subsequently prioritisation of management actions has been problematic. The recent growth in the use of Bayesian network tools [see Henderson *et al.*, 2008] for ecological risk assessments has resulted in major advances in better understanding and managing ecosystems despite their inherent complexity.

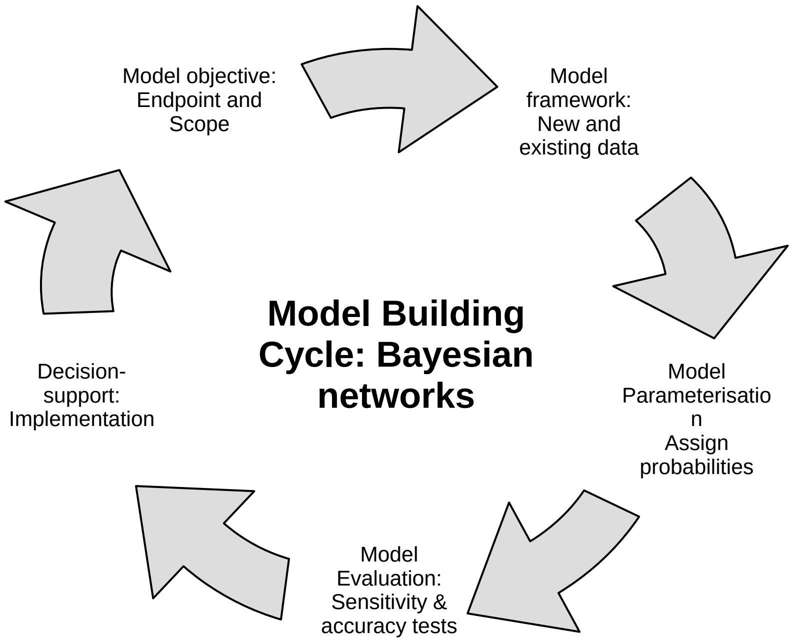

Figure 2: Steps followed in building a Bayesian network.

Abductive inference and its quantification by means of Bayes's theorem can also further reduce bias and provide a framework for the efficient accumulation and use of evidence [Newman & Evans, 2002]. Via abductive inference, Bayesian networks also offer a means of examining attributable risk or probability of causes, and for combining data sources (e.g. experiential and non-experiential) to yield information that neither study alone can do [Pearl, 2000]. Tools and techniques (including applying weightings) for combining information sources have been trialled and reviewed elsewhere [e.g. Pollino & Hart, 2006; Pollino *et al*., 2007].

At the core of an ecological risk assessment is the assessment of causality [Newman & Evans, 2002]. Bayesian networks can be used to formulise quantification of risk, representing likelihood/hazards as system variables that are linked to consequences/effects (such as environmental values that can act as model endpoints) within a risk assessment framework. As such, Bayesian networks readily enable relationships between multiple stressor/hazards and multiple endpoints to be modeled in a holistic way. The steps taken in building a BN facilitate the ERA process and provide a tangible outcome that can be used to examine and prioritise key risks, including uncertainties. As both processes are iterative, they also fit within an adaptive management context. BNs can also be extended for risk management planning and decision-making, where other non-scientific factors (e.g. cultural, economic, etc.) can be included in models.

Two case studies where Bayesian networks were developed in a risk assessment context are briefly described below. For an in-depth description of the case study methods used for expert elicitation and model parameterisation and evaluation, please refer to our previous papers [Pollino & Hart, 2006, Pollino *et al*., 2007].

2.1 Case Study 1: Goulburn Catchment, Australia

An ecological risk assessment was undertaken in the Goulburn Catchment (Victoria, Australia). Water from the Goulburn River, and its tributaries, have a number of uses, including: irrigation, native and trout fishery, recreational boating and town supply.

In the problem formulation phase, stakeholders identified the native fish abundance and diversity in the Catchment as an asset under threat. A risk assessment was designed to support future decision-making in the catchment, by:

- Quantifying linkages between threats and native fish communities throughout the catchment, considering both site and reach scales;
- Prioirtising risks to native fish communities;
- Communicating uncertainties in predictions and identifying key knowledge gaps.

A conceptual model of the catchment was constructed, linking system variables and endpoints. This was developed in association with catchment managers and ecologists, and formed the structure of the BN. Model endpoints were Future Abundance and Future Diversity of native fish. The model consists of five interacting components, water quality, hydraulic habitat, structural habitat, biological potential and species diversity [Pollino *et al.*, 2007].

A BN was constructed to represent the 23 sites in the Goulburn Catchment, which were aggregated into 7 priority reaches. Prediction can be made per site or reach, or as aggregates. Model parameters were derived in 3 ways: expert only; data only; and a combination of expert and data. Data parameterization used the expectation maximisation (EM) algorithm, where missing data was predominately due to an absence of monitoring data. For details of model parameterisation and evaluation, see [Pollino *et al.*, 2007]. The relationships in the BN had not been characterized previously.

Despite the high uncertainties associated with the lack of knowledge of the relationships between variables and lack of available data, the BN developed has the ability to predict the abundance and diversity of native fish communities based on existing and predicted changes to environmental conditions. Consequently, it also has the capability to assist in determining what management options are most favourable for maintaining and rehabilitating fish communities at multiple spatial scales. However, it is recognised that the model requires further testing in the field to determine its accuracy pre- and post-management interventions or system changes.

The Bayesian network endpoints are directly measurable and results are easily interpreted by stakeholder groups, including system managers and ecologists. Management scenarios or system changes can be investigated to inform decision-making, where top-down and bottom-up reasoning can be applied. As model predictions are likelihoods, they are directly applicable to risk management. Sensitivity analyses allowed prioritisation of threats to environmental value, and key data and knowledge gaps identified [Pollino *et al.*, 2007], allowing areas needing further research to be identified and prioritised.

2.2 Case Study 2: Ok Tedi-Fly River, Papua New Guinea

The mining industry has a long legacy of causing major environmental impacts. Although most mining activities today have improved environmental operations, unacceptable changes to the environment still occur. Increasingly, the mining industry is using risk-based approaches to better predict and manage unacceptable changes to downstream ecosystems.

The Ok Tedi copper and gold mine, operated by Ok Tedi Mining Limited (OTML), is located in the mountains of Western Province of Papua New Guinea. The mine has been operating since 1984 and discharges approximately 164,000 tonnes of waste rock and 82,000 tonnes of tailings per day to the Ok Tedi, with these mine-derived sediments being transported downstream to the Fly River. As a result, channel aggradation, up to 70 m in the upper sections of the catchment, has occurred. Studies investigating the impacts of mine-derived materials in the Ok Tedi-Fly River system have raised concerns that acid rock drainage (ARD) and the associated liberation of increased concentrations of heavy metals, in particular copper, may cause additional environmental damage.

In order to assist decision-making regarding future mine operations, OTML decided to undertake an ecological risk assessment, with the aim:

- To assess the risks to environmental values in the Ok Tedi and Fly River system from ARD; and
- To test the effectiveness in mitigating these risks by removing 85% sulphur from the tailings before they are discharged to the river system.

Assessment endpoints in this study were: drinking water standards; the availability and edibility of fish; the access to forest products and availability and edibility of terrestrial foods; and primary productivity (focussing on algae). Five Bayesian network models were constructed [Pollino & Hart, 2005; Pollino & Hart, accepted].

The model causal structures were based on conceptual models developed by mine managers, ecologists, chemists, geologists and ecologists. BN models represent two major riverine processes, water chemistry and sediment processes. Water chemistry is under the influence of ARD processes (ARD and no ARD scenarios can be tested). Channel aggradation is influenced by sediment transport. The outcomes of sub-models are integrated into endpoint variables. Activities at the mine and their subsequent influence on system processes and model endpoints are the focus of management activities in models. The Bayesian network represented 5 reaches over a number of time intervals.

Two physical models were used to parameterise water quality and channel aggradation processes. The physical models used were OkChem-OkARD [EGi, 2005], a purpose build water quality model and HEC-6 [Pickup & Cui, 2003]. Expert opinion and/or research and monitoring data were used to parameterise ecological endpoints. Sensitivity analysis was conducted on all models and predictive accuracy tests were used, where possible. For details of model parameterisation and evaluation, see [Pollino *et al.*, 2007] and Pollino *et al.* [accepted]. The models were used in risk management processes, where a series of mine operation scenarios, and the resultant ecological outcomes, were investigated. The findings were used to assist in scenario-based risk management and to guide investments in improved mining operations and further research.

3. CONCLUSION

As embodied in many existing regulatory guidelines [e.g. USEPA, 1998; ANZECC, 2000], the ecological risk assessment framework has been advocated as a 'reductionalist' approach to management. This type of approach has been highly criticised as not being representative or applicable to ecosystems and the needs of natural resource mangement. To broaden the relevance of risk analysis techniques to ecosystems, a 'holistic' approach to risk assessment is now advocated [Assmuth & Hilden, 2008; Hart *et al.*, 2007; Macleod *et al.*, 2007], where multiple stressors and their interactions across an ecosystem are taken into account [Leuven & Poudevigne, 2002]. To achieve this, a nested hierarchical framework is often used to break down complex ecosystem processes into simpler levels of ecological organisation, which can be more easily reassembled using a bottom up process [Leuven & Poudevigne, 2002; Levin, 1999].

This bottom up process for a risk assessment is widely advocated in modern risk assessment frameworks [Hart *et al.*, 2007]. Bayesian networks can also be used to represent and integrate knowledge of processes, or existing models, from other fields (e.g. hydrology, sediment transport and contaminant life cycle models), and consequently, they are not intended to 'reinvent the wheel'. As models are iterative, this knowledge can be regularly updated, fitting into an adaptive management framework of continual learning. Wider application of Bayesian methods can reduce problems associated with causality assessments, reduce conflicts emerging from less formal integration of available evidence, and more effectively use limited resources for risk assessments, and their use in risk management [Newman & Evans, 2002].

Finally, BNs are also ideal for scenario based risk management, where scenario analysis is a key element required for assessing risk [Hart *et al.*, 2007]. The BNs allowed for greater analysis of a decision, where uncertainty was a key component of the analysis, and risk management outcomes, where decisions are liable to be open to scrutiny.

REFERENCES

Assmuth, T. & Hilden, M. (2008) The significance of information frameworks in integrated risk assessment and management. *Environmental Science & Policy*, **11**, 71-86.

Borsuk, M.E., Stow, C.A., & Reckhow, K. (2004) A Bayesian network of eutrophication models for synthesis, prediction and uncertainty analysis. *Ecol. Modelling*, **173**, 219-239.

Burgman, M. (2005) *Risks and decisions for conservation and environmental management* Cambridge University Press, Melbourne.

EPA (1998). Guidelines for Ecological Risk Assessment. United States Environmental Protection Agency, Washington D.C.

Hart, B., Pollino, C., White, A., Grace, M., Burgman, M., Fox, D., Carey, J., Chee, Y.E., Henderson, B., & Bui, E. (2007). Risk-based approaches for managing contaminants in catchments. In *Salt, Nutrient, Sediment and Interactions: Findings from the National River Contaminants Program* (eds S. Lovett, P. Price & B. Edgar), pp. 127 - 140. Land and Water Australia, Canberra.

Henderson, C., Pollino, C.A. and Hart, B.T. (2008). Bayesian modelling as a basis for risk-based Environmental Flow assessments: Literature Review. Report by iCAM, ANU and Water Science Pty Ltd for the National Water Commission, Canberra

Hubbard, D. (2007) *How to Measure Anything: Finding the Value of Intangibles in Business.* John Wiley and Sons, New York.

Leuven, R.S. & Poudevigne, I. (2002) Riverine landscape dynamics and ecological risk assessment. *Freshwater Biol.*, **47**, 845-865.

Levin, S.A. (1999) Towards a science of ecological management. *Conservation Ecology*, **3**, 6-9.

Macleod, C.J.A., Scholefield, D., & Haygarth, P.M. (2007) Integration for sustainable catchment management. *Science of the Total Environment*, **373**, 591-602.

Newman, M.C. & Evans, D.A. (2002). Enhancing Belief during Causality Assessments: Cognitive Idols or Bayes's Theorem. In *Overview of Ecological Risk Assessment in Coastal and Estuarine Environments* (eds J. M.H. Roberts, M.C. Newman & R.C. Hale), pp. 73 - 96. Lewis Publishers, New York, New York, USA.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference.* Cambridge University Press, Cambridge, UK.

Pollino, C.A. & Hart, B.T. (2005). Assessment of the risks to the environment and human health from mine-derived heavy metals in the Ok Tedi and Fly Rivers, PNG. Unpublished report by Water Science Pty. Ltd. and Water Studies Centre, Monash University for Ok Tedi Mining Limited dated August 2005.

Pollino, C.A. & Hart, B.T. (2006) Bayesian decision networks - going beyond expert elicitation for parameterisation and evaluation of ecological endpoints. In Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modelling and Software Society (eds A. Voinov, A. Jakeman & A. Rizzoli), Burlington, USA, July 2006.

Pollino, C.A. & Hart, B.T. (accepted) Assessing the ecological risks of a mine operation in Papua New Guinea: Part 1. Background and Methodology. *Risk Analysis*.

Pollino, C.A., Woodberry, O., Nicholson, A.E., Korb, K.B., & Hart, B.T. (2007) Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software*, **22**, 1140-1152.

Stow, C.A. & Borsuk, M.E. (2003) Enhancing causal assessment of fishkills using graphical models. *Ecosystems*, **6**, 11-19.