Ecological Modelling 358 (2017) 1–9

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Common quandaries and their practical solutions in Bayesian network modeling

Bruce G. Marcot

U.S. Forest Service, Pacific Northwest Research Station, 620 S.W. Main Street, Suite 400, Portland, OR, USA

ARTICLE INFO

Article history:
Received 21 December 2016
Received in revised form 12 May 2017
Accepted 13 May 2017
Available online 24 May 2017

Keywords:
Bayesian networks
Modeling problems
Modeling solutions
Bias
Machine learning
Expert knowledge

ABSTRACT

Use and popularity of Bayesian network (BN) modeling has greatly expanded in recent years, but many common problems remain. Here, I summarize key problems in BN model construction and interpretation, along with suggested practical solutions. Problems in BN model construction include parameterizing probability values, variable definition, complex network structures, latent and confounding variables, outlier expert judgments, variable correlation, model peer review, tests of calibration and validation, model overfitting, and modeling wicked problems. Problems in BN model interpretation include objective creep, misconstruing variable influence, conflating correlation with causation, conflating proportion and expectation with probability, and using expert opinion. Solutions are offered for each problem and researchers are urged to innovate and share further solutions.

1. Introduction

Bayesian network (BN) models are essentially graphs of variables depicted and linked by probabilities (Koski and Noble, 2011). BNs have become a favorite framework for assessment and modeling in a wide variety of fields, ranging from evaluating effects of climate change (Moe et al., 2016), trade-offs among ecosystem services (Landuyt et al., 2016), management of coastal resources (Hoshino et al., 2016), and many other applications (Pourret et al., 2008).

BN modeling is attractive because it is relatively easy to do, it provides a highly intuitive and interactive interface, and it is easy to manipulate to test model assumptions and scenarios. However, because of its growing popularity and ease of construction, BN modeling carries risks when misapplied. Further, a number of unobvious errors and problems can arise from model development through model interpretation, solutions for which few publications address.

Textbooks on BN modeling tend to not delve into many practical modeling problems and solutions, including books on introductions to BN modeling (Koski and Noble, 2011; Hobbs and Hooten, 2015; Holmes and Jain, 2010; Darwiche, 2009; Kjaerulff and Madsen, 2007; Neopolitan, 2003), and use in decision analysis and risk assessment (Jensen and Nielsen, 2007; Fenton and Neil, 2012; Condamin et al., 2006), artificial intelligence programming (Korb and Nicholson, 2010a, 2010b), causal modeling (Sloman, 2009), and other fields. Although the number of generally accessible journal articles on BN modeling has continued to increase in recent years, achieving an exponential growth at least during the period from 1980 to 2000 (Fig. 1), they too seldom offer practical solutions to many common problems. Other BN modeling references and guidelines have been produced that may provide some solutions, but many appear as unpublished or largely inaccessible reports (e.g., Bromley 2005), gray literature, government agency publications (e.g., Amstrup et al., 2008; Das, 2000; Pollino and Henderson, 2010), university reports (e.g., Korb et al., 2005), application user guides (Conrady and Jouffe, 2015; Woodberry and Mascaro, 2014), online tutorials (e.g., www.cs.ubc.ca/~murphyk/Bayes/bnintro.html), and embedded and online help systems of specific BN modeling programs (e.g., www.norsys.com/WebHelp/NETICA.htm).

A fine exception is the more generally available book chapter by Korb and Nicholson (2010a, 2010b) who covered a number of BN modeling mistakes pertaining to problem statements, model structures, and model parameters. The current paper is intended to complement and expand upon their highly useful contribution and other sources cited above, and to call for clarity in definitions of terms used in BN modeling. My intent is not to critique specific, published models, but rather to summarize major, common problems I have encountered in BN model construction and interpretation, and to offer practical solutions. I draw mostly from my >25 years' experience with working in this modeling construct (e.g., Marcot, 1990, 1991, 2007, 2012; Marcot et al., 2001, 2006), learning from many fine colleagues and collaborators, and advising others on BN modeling projects.

E-mail address: bmarcot@fs.fed.us

http://dx.doi.org/10.1016/j.ecolmodel.2017.05.011
0304-3800/Published by Elsevier B.V.

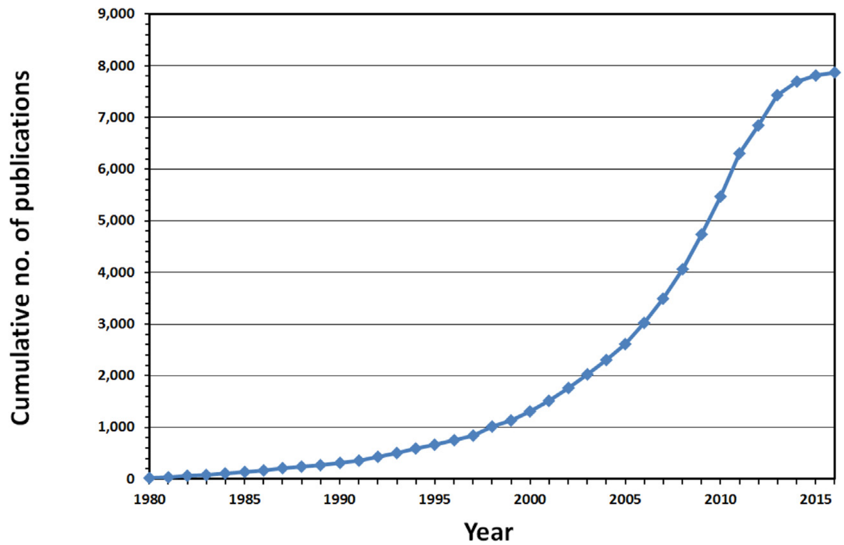

Fig. 1. Cumulative number of publications (journal articles, books, and reviews) on Bayesian network modeling, 1980–2016, based on items cataloged by JSTOR (www.jstor.org). Keyword searched: Bayesian network. JSTOR accessed 12 May 2017.

2. Quandaries and solutions in Bayesian network construction

BN models are constructed by specifying variables with continuous or, more commonly, discrete states, and linking them with conditional probability values that specify the immediate relationship between and among the variables. This section addresses problems commonly encountered in BN modeling pertaining to construction of the network and setting the probability values underlying each variable. Probability values typically appear in BN models (1) as an unconditional (or prior) probability table for variables (nodes) that are dependent on no other variables, that is, that have no "parent" nodes, or (2) as a conditional probability table (CPT) for variables that are dependent on other variables. Probabilities are typically depicted on [0,1] scales.

Issue: Thinking that every CPT must contain values that span [0,1] (i.e., 0% and 100%).

Why this is a problem: When specifying CPT values manually – rather than from a machine-learning algorithm using a data set – some modelers start by denoting worst and best combinations of inputs with values of 0 and 1, respectively, and then gauging intermediate values for the rest of the table, as a short cut method. However, forcing CPTs to always "peg corner values" of 0 and 1 probabilities (e.g., Table 1) can unduly affect the sensitivity structure of a model by overstating the influence of a variable; i.e., an incremental change of probability from 0.99 to 1 may have a far greater influence than a change from 0.50 to 0.51. This problem may arise when specifying values from expert knowledge instead of via machine-learning algorithms.

Practical solution: Depending on the objectives of the model and how CPTs are formulated – i.e., induced from data or proposed from expert knowledge – CPT values can be interpreted as representing frequencies of occurrence of outcomes given conditions. As such, values of a CPT would not necessarily need to, and often should not, span [0,1].

Issue: Vague node names and unmeasurable node states.

Why this is a problem: BN models developed from expert knowledge – particularly from multiple experts – often contain nodes that are ill-defined and unmeasurable, making the overall model little more than an unverifiable belief system, thus reducing its potential credibility, practical application, and validation (Pitchforth and Mengersen, 2013). An example could be in a model evaluating habitat conditions for a wildlife species that contains a node named "population response," where the states of the node might be given as 'high' and "low." In this example (Fig. 2), if the node name is not defined with a specific unit of measure, it could be interpreted as many possible measures of response such as percent occurrence of the species, population density, population trend, a resource selection function, and so on. The states as given would be ambiguous and unclear if they do not each have clear cutoff conditions or values, leading different observers to interpret and measure them in very different ways, resulting in different model outcomes for the same conditions.

Practical solution: As far as possible, all nodes and their states – particularly input nodes that drive the model – should be clearly defined and empirically measurable, whether categorizing, counting, or quantifying, prior to populating the probability tables. That is, it should be possible to send someone into the field, into a lab, etc., and know exactly and unambiguously what the node represents and consistently how to measure it.

Issue: Too many parent nodes and no higher thinking about summary variables.

Why this is a problem: This can be problematic for models particularly in which CPTs are to be populated by expert knowledge, rather than induced from a data set and a machine-learning algorithm. Too many parent nodes – along with too many discrete or discretized states of parents and child nodes – result in massive, multi-dimensional CPT structures that are difficult to impossible to visualize and for which to effectively specify values (e.g., Fig. 2a). In a way, this throws all affector variables into the same pot with little thinking about relative causal influences that could better be partitioned out (Fig. 2b).

Practical solution: Combine some of the parent nodes, a few at a time, into fewer intermediate nodes (although this in turn can run the risk of the previous issue of vague node names and states). This should be done based on logical combinations, causal structures, and consideration of potential latent variables not otherwise explicitly expressed in the influence diagram. Then, the fewer inter-

Table 1
Conditional probability table of node 'Maternal Roosts' in Fig. 2b, illustrating probability values "pegging the corners" to span [0,1]. Outcome values in each row represent probabilities given a specific prior condition, whereas outcome values in each column represent likelihoods of various prior conditions given a specific outcome. In the first row, the odds ratio of Adequate:Inadequate is 0.8:0.2 or 4:1. In the Adequate outcome column, the risk ratio (or likelihood ratio) of prior conditions [Present, Present, ≥30] and [Present, Present, 20 to 30] is 0.8:1.0.

Parent nodes			Maternal Roosts outcome	
Caves or Mines	Large Snags/Trees	Summer roost cave temp (C)	Adequate	Inadequate
Present	Present	≥30	0.8	0.2
Present	Present	20 to 30	1.0	0.0
Present	Present	<20	0.8	0.2
Present	Absent	≥30	0.1	0.9
Present	Absent	20 to 30	1.0	0.0
Present	Absent	<20	0.1	0.9
Absent	Present	≥30	0.7	0.3
Absent	Present	20 to 30	0.7	0.3
Absent	Present	<20	0.7	0.3
Absent	Absent	≥30	0.0	1.0
Absent	Absent	20 to 30	0.0	1.0
Absent	Absent	<20	0.0	1.0

mediate nodes can be combined into the original child node so that it has a far smaller and more tractable CPT dimensionality.

Consider first exploring covariate correlations, such as with simple bivariate Pearson correlation coefficients (r), to exclude highly correlated variables with $|r| \geq 0.7$. Also consider exploring path regression or general linear models first to determine partial correlation strengths to prioritize key covariates and possibly exclude

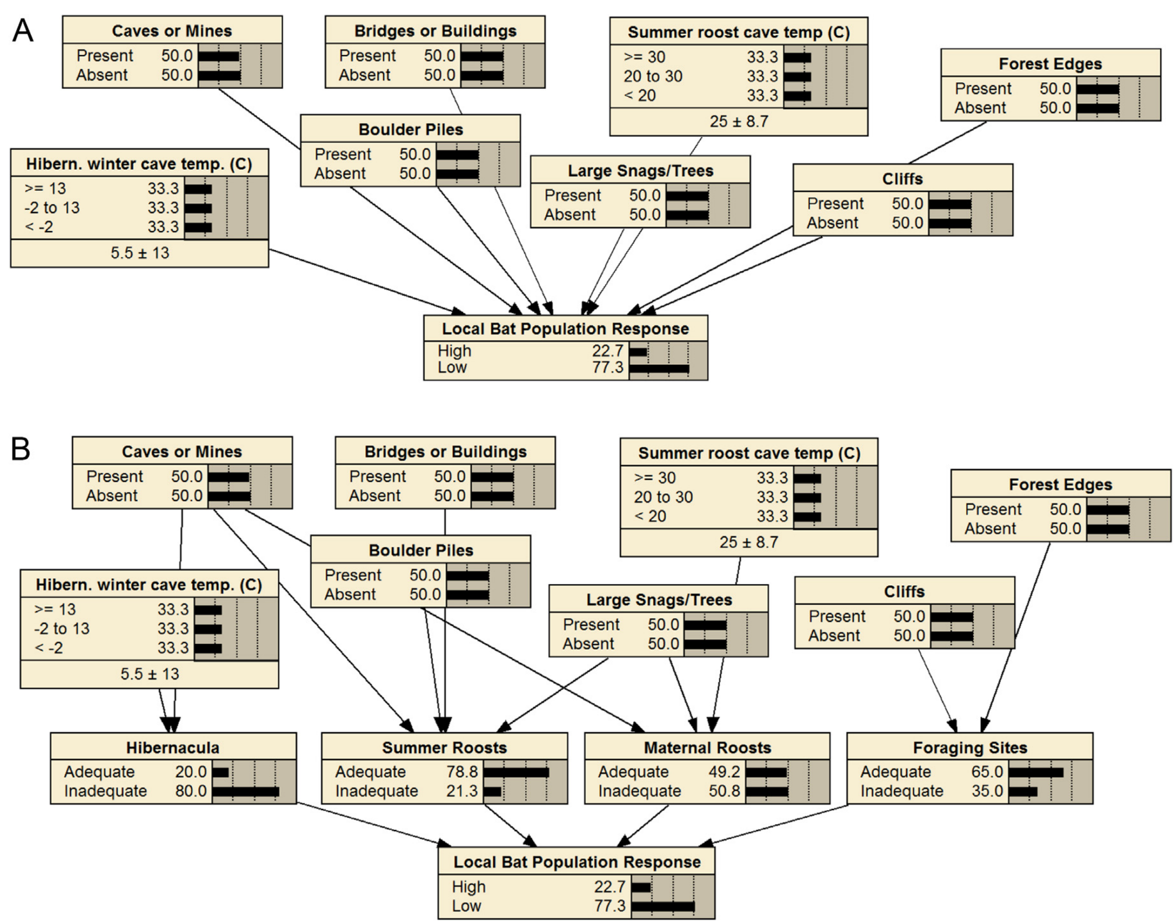

Fig. 2. Structuring a Bayesian network model with intermediate nodes so as to make conditional probability tables smaller and tractable for parameterization by expert elicitation. The model evaluates the potential response of Townsend's big-eared bats (*Corynorhinus townsendii*) to conditions pertaining to their hibernacula, roost, and foraging sites (after Marcot et al., 2001).
(a) Initial expert-guided model structure with all covariates directly linked to the bat response outcome variable, the conditional probability table of which contains 1152 values.
(b) The same model with four summary nodes inserted to improve the logical structure and to greatly reduce the size of the conditional probability table of the outcome variable to only 32 values.

minor ones. This will simplify the model and reduce the size of the probability tables.

Issue: Ignoring the role of potential latent variables in expert-structured networks.

Why this is a problem: A latent variable is one inferred from the mathematical relation among other observed variables. Not allowing for those relations can greatly reduce the classification accuracy and explanatory power of a model.

Practical solution: Provide links between predictor variables if they are related causally, if they show significant correlation, or if they are otherwise dynamically related. In this way, a parameterized network then can treat such node cliques as an collective, emergent latent influence, such as is done with latent variables in structural equation modeling (Arhonditsis et al., 2006; Capmourteres and Anand, 2016; Grace et al., 2010).

Issue: Not considering confounding variables when constructing a causal network.

Why this is a problem: Sometimes used synonymously with "lurking variable," a confounding variable is one that is not considered or measured in a study but that could influence other variables. Not including confounding variables can attribute causality to the wrong predictor variable, can bias sensitivity analyses of a model, or at best, can fail to account for fuller explanatory power.

Practical solution: One can provide at least a "ghost node" placeholder for suspected confounding variables, to be specified, adjusted, and parameterized upon further data collection and study. A ghost node – defined here as one with no influence – would be included in a network such that would have no effect on its child nodes whose probability distributions would not vary by the ghost node's different states. But it serves at least as a marker to explicitly denote recognition of a higher-order influence. Other solutions include randomizing the effect of a confounding variable over other variables (Guillemette and Larsen, 2002), or factoring out the variable if it contributes to autocorrelation (King et al., 2005).

Issue: Discounting or eliminating one expert's input (e.g., CPT values) if it diverges from other experts' inputs, because it is seen as an outlier opinion.

Why this is a problem: The one lone-wolf expert might have unique knowledge or experience of a situation that other experts do not. Ignoring that knowledge could bias the resulting model and make it less robust to new situations.

Practical solution: Poll experts individually in structured deliberations (Gregory et al., 2012), if >1 expert is used to structure or parameterize a BN model, whereas using an expert paneling approach to reach consensus, such as with the Delphi method (Hsu and Sandford 2007; Delbecq et al., 1986), masks individual variations. Have them identify the strength of evidence as well as key uncertainties for their responses and inputs. Also define up front if the objective is to create one model to represent all experts' knowledge collectively, or to create >1 model so as to represent the potential differences among the experts; either way, do not summarily discount the lone wolf without valid rationale.

Another solution is to explicitly include a node in the network to represent each expert, either by name or code if they are to remain anonymous, and link that node to whichever other variables that they directly influence (Uusitalo et al., 2005). This structure can be useful for directly quantifying the sensitivity of model outcomes to experts' inputs, and can provide the option to weight the experts if needed.

Issue: Not linking parentless input nodes that are themselves correlated, nor even checking for their correlation to begin with.

Why this is a problem: In Bayesian networks using Dirichlet-multinomial prior probabilities (Spiegelhalter et al., 1993), it is assumed that unlinked parentless (input) variables are uncorrelated. Failing to denote correlated variables with links in a Bayesian network model incorrectly represents the probability structure of the network, which could bias the sensitivity and the influence dynamics of the model and validation outcomes.

Practical solution: If empirical data sets are available, run simple bivariate Pearson correlations on the input variables and, if they are all retained in the model, link the ones with high correlation (Stevens and Thadani, 2006), particularly when $|r| \geq 0.7$ or so, if they are to be retained in the network. If empirical data sets are unavailable, consult with experts to determine which input variables are likely to be correlated in their experience, and link them accordingly. Link them according to logical causal relations if one exists (linking affector causal agents to resulting effector agents); otherwise, use as the parent node that variable in a correlated pair that is easier or less costly to determine or observe. If, in running the model, the values of both members of a correlated pair of variables will be specified, then it generally doesn't matter how they are linked. Also, when prior knowledge exists about correlations among variables, a more robust model structure may result from applying supervised techniques to identify and designate link structures around known interdependencies (Uusitalo 2007).

Issue: No peer review of the model with reconciliation.

Why this is a problem: This can be a problem of scientific credibility, particularly with expert-structured networks.

Practical solution: Especially for expert-based networks, seek peer review at two stages of model development: (1) specification of the initial influence structure showing node names, states, and their linkages, along with clear definitions of the node names and each state, including potential units of measure and means of measurement of their states; (2) the fully-operational alpha-level model including its CPT values, equations used in nodes, and other aspects of the model parameters (Marcot et al., 2006). Review of the initial influence structure is sometimes overlooked but it can help modelers avoid investing in a poorly-structured network from which later changes become increasingly difficult.

Issue: No specific tests of model calibration and especially validation.

Why this is a problem: Testing a model against a set of data which may have been used in some way to build the model is a measure of model *calibration*. Testing a model against a set of independent data not used to build the model is a test of model *validation* (e.g,. Caley et al., 2014). Depending on the purpose of the model, not conducting tests of model calibration and validation can result in a tool with no clear determination of its accuracy and expected error rates (Type I and Type II). This is not so much a problem with models intended to basically represent the state of knowledge, or models for which case data are simply unavailable, impractical, or nonmeaningful, but can be problematic for other models aimed at prediction, projection, and causal explanation.

Practical solution: Where data are available or can be gathered, compile a case file of known conditions and their outcomes as means to test model prediction (classification) accuracy. Many metrics are available to denote model prediction or classification accuracy and error rates (Marcot, 2012); select the metric(s) that best fits the purpose of the model. Park and Budescu (2015) also suggested a method of aggregating probability intervals to improve calibration. Of increasing popularity in validation testing is use of k-fold cross-validation or the leave-one-out technique (Arlot and Celisse, 2010; Forio et al., 2015).

Where data are unavailable or cannot be gathered, it may be possible to query other experts and compile a case file of their experience and known or predicted outcomes, and use that to validate the model. However, such results would then be more appropriately interpreted as verification of the model against outcomes provided by a set of independent experts' knowledge.

Issue: Confusing calibration with validation.

Why this is a problem: Following the above issue, calibration and validation, as defined above, differ in their objectives. Confusing calibration with validation leads to incorrectly overstating model accuracy. Calibration tests how well a model is constructed to represent the conditions and variations in a given data set, whereas validation tests how well the model can predict (or classify) independent conditions using data that were not applied to its construction. When a model has a high degree of fit to a set of data, calibration accuracy will likewise be high (issues of overfitting aside), but this does not necessarily mean that the model is useful for other conditions and contexts beyond those of the data set.

Practical solution: Conduct separate tests of model calibration and, where possible, validation.

Issue: Overfitting of a machine-learned network structure or probability parameters to a data set.

Why this is a problem: Overfitting classically results in a model with no degrees of freedom; in Bayesian networks, this means that the model has very limited robustness for application to situations outside those in the data used to structure or parameterize the model (Kuschner et al., 2010).

Practical solution: If a data set is large enough, parameterize the model using a bootstrapping technique (e.g., using leave-one-out), or otherwise conduct k-fold cross-validation. If the data set is very small compared to model complexity (numbers of nodes, states, and links), consider first parameterizing it using expert knowledge, and then test and update it with the data set, instead of inducing the model structure and parameters strictly from the data. Measures of, and means of avoiding, overfitting in BN models are also provided by Cawley and Talbot (2007), Han et al. (2014), and Jakulin and Rish (2006). Pollino et al. (2007) developed BN models for ecological risk assessment by combining data-driven and expert-specified approaches, which could help alleviate overfitting.

Issue: Models parameterized from machine-learning probability algorithms often have many unspecified or extreme values in the resulting probability tables.

Why this is a problem: Inducing (machine-learning) probability structures from a case file by incorporating the case file or using a gradient or expectation maximization (EM) algorithm (e.g., Dlamini, 2011) often results in many unspecified values or "holes" in the CPT of the outcome variable (node). The holes are typically filled with uniform probability distributions (automatically in Netica®, Norsys Inc.). This problem occurs very commonly because typical empirical case files often lack all possible combinations of all states of input (parent) nodes, especially in sample sizes large enough by which to derive frequency distributions of outcomes; and if only one case is available for a specific combination of input states, the resulting probabilities will be skewed as [1,0,0.]. It is also more apt to occur with networks having more input nodes and states, fewer examples in the case file, or both.

The resulting model then really pertains only to the conditions represented in the case file, that is, only to the input combinations for which there were data to train the model. For all other conditions not represented in the case file, the model presumes or defaults to uniform probability outcomes, which represents total uncertainty. Further, such a condition will skew the model's sensitivity structure and scenario run results.

Practical solution: This problem is not trivial. First, the conditional probability tables developed from data-driven algorithms can be quickly scanned to determine the specific occurrence and severity of this condition, by locating the uniform probability distributions. Then, various approaches can help reduce this problem:

(1) Initially structure the model (establishing the linkages among nodes) using expert guidance to deliberately (a) reduce the number of input variables, (b) include intermediate summary variables, and (c) reduce the number of states in input and outcome variables. This should greatly decrease the complexity and size of the output node CPT, but still will not guarantee that all combinations of inputs are represented as cases in the case file. In one approach, Myllymäki et al. (2002) recommended methods which use ecologically significant breakpoints and that minimize the number of states so that each interval contains enough data to parameterize and run the model.

Alternatively, CPT size can be kept small by using a naïve Bayes approach or a variant thereof (e.g., tree augmented network, Jiang et al., 2012) to structure a network from a data set. Naïve Bayes structures result in links directed from a response variable to their covariates (affector variables). This can be confusing to interpret because superficially it appears that cause and effect relationships seem reversed. However, after a naïve BN model is defined and parameterized with CPT values, the direction of node links can then be reversed if a causal structure is desired, and this does not alter the functioning of the model.

(2) Restructure the model, and/or expand the case file, so that all combinations of inputs are represented in the case file by »1 case; this may take greatly simplifying a first-generation "alpha" level (Marcot et al., 2006) model to better match a given case file, or securing additional case examples to fill in expected holes.
(3) Alternatively, initially populate the probability tables (especially the CPTs) in the model with best expert knowledge, and then when updating the model from a case file, revise but do not delete those probability values. In this way, the "holes" in the outcome CPT will retain at least the expert-generated values, although thought needs to go into the relative weighting of expert values versus data cases. Then, you will still want to conduct a series of structured influence runs (Marcot, 2012) to determine if incorporating the case file resulted in any non sequitur outcome patterns. Using this approach, initially structure the model from expert knowledge rather than using a machine-learning approach which can result in broad, shallow networks with large conditional probability table structures that may be very difficult to interpret and manually parameterize.
(4) Other approaches to filling in the "missing" (uniform probability) outcomes can include exporting the CPT to a spreadsheet and then create bivariate line plots and using an interpolation (doesn't have to be linear!) or imputation approach to smooth the probability values. An imputation approach might use multivariate statistics such as canonical correlation to determine

the value of a missing probability based on where it falls in a state-space relative to other known values and their inputs. With complex CPTs, such as with >3 parent variables, interpolation can become difficult to logically interpret. Alternatively, Oniśko et al. (2001) proposed a method employing Noisy-OR gates that reduces dependence on size of the case file to fill in missing probabilities. This algorithm has been incorporated into the GeNIe Modeler® (BayesFusion, LLC) BN modeling program.

(5) Live with the holes! Perhaps they are not critical to subsequent use of the model for particular locations or contexts. Or they could even be useful representations of data gaps and incomplete knowledge (from the case file). Or, fill in only those CTP holes for which you know you will need for new situations, and let the rest remain as uniform probabilities.

Issue: Trying to model a truly "wicked problem."

Why this is a problem: Wicked problems, by definition, are intractable and insolvable (Rittel and Webber, 1973).

Practical solution: Deconstruct the problem definition into more elemental component problem statements that can be modeled (Sethi and Dalton, 2012). Apply expert judgment – here, defined as the process of applying personal experience in weighing evidence or information to reach a conclusion – to components of the problem (Richards et al., 2014). Wicked problems sometimes can be made tractable by use of collective experience of multiple experts (van Bueren et al., 2003; Conklin, 2005), as in well-orchestrated expert panels (e.g., Orsi et al., 2011).

3. Quandaries and solutions in Bayesian network interpretation

In this section I discuss problems and solutions pertaining to the interpretation of using a Bayesian network for a stated purpose.

Issue: Unclear initial objectives, leading to "objective creep."

Why this is a problem: Unclear modeling objectives typically lead to poorly-designed models that are expected to provide a wide range of services such as diagnosis, prediction, forecasting, scenario assessment, mitigation, and so on (Table 2). No one model can do it all. As model development continues, the expectations for its use and the kinds and numbers of variables incorporated often tend to "creep," meaning constant change or expansion. This results in a poorly-considered model structure with unclear utility and often a waste of effort and money in its development, as well as a model with low to no scientific credibility.

Practical solution: Begin any modeling project with a clear, simple, written statement of the purpose and objective (Chen and Pollino, 2012); select one main objective such as those presented in Table 2. State this as a problem to be solved, not as a model structure per se. Remember that, for most practical modeling projects, "building a model" is a method, not an objective.

Issue: Conflating probability, likelihood, odds, and risk.

Why this is a problem: Confusion over statistical terms can lead to misinterpretation of the purpose, interpretation, and application of a BN model. Using the terms interchangeably is a common error, whereas they each serve very different purposes.

Practical solution: Define and use these terms with clarity (Table 1). *Probability* refers to possible outcomes given a set of specified initial conditions, and is depicted as rows in a conditional probability table wherein all outcome probability values sum to 1. *Likelihood* refers to possible initial conditions that could lead to a specified outcome, and is depicted as columns in a conditional probability table, wherein the sum of column values do not necessarily sum to 1; normalizing the values in a column approximates a normalized likelihood function. *Odds* (or *odds ratio*) refers to the probability of one outcome compared with another outcome (Pearl 2000); e.g., the odds of outcome A compared with outcome B is the posterior probability of outcome A divided by that of outcome B. In addition, *risk ratio* (or *likelihood ratio*) is the likelihood of one initial condition given an outcome, divided by the likelihood of another initial condition given that same outcome (Rothman and Greenland 1998). And in general, *risk* is the probability of an outcome weighted by its utility; *risk* often is used synonymously and erroneously with just probability.

In particular, when parameterizing or peer-reviewing a conditional probability table, especially when constructed from expert knowledge, it can be quite useful to check both probability (rows) and likelihood (column) values, including selected odds ratios and risk ratios, to determine if the model would conform to expectations.

Issue: Conflating conditional probabilities with confidence in their veracity.

Why this is a problem: Conditional probabilities denote potential outcomes given a set of conditions. Although uncertainty in the resulting outcomes can be expressed by "feathering" probability values across >1 outcome, this is not necessarily the same as denoting the degree of confidence one should place in those values, as one could be highly confident of well-distributed (platykurtic) probabilities, or have low confidence in highly modal (leptokurtic) probability distributions. Current BN modeling programs provide limited to no capacity for explicitly denoting confidence in probability values, such as were incorporated into expert system programs of the 1980s and 1990s. Lacking knowledge of the confidence behind probability values might overstate their veracity (Qin et al., 2011). Assessing confidence and validity is particularly problematic with BN models constructed from expert elicitation (Pitchforth and Mengersen, 2013).

Practical solution: This is an interesting research problem in BN modeling that has been addressed in at least two very different ways. Van Allen et al. (2008) proposed denoting the uncertainty of BN random variables with error bars, whereas Wilhere (2012) developed a method to discretize BN variables so as to incorporate numerical error, and to incorporate measurement error in the calculation of expected values of model outcomes. Some programs, such as Netica® and GeNIe®, provide a measure of "experience" calculated as equivalent sample sizes of relevant cases that contribute to probability values; this is useful when parameterizing probability values via machine-learning algorithms and case files, but does not address levels of confidence that an expert might place in their own specified probabilities.

Issue: Assuming that outcomes are always most sensitive to the closest nodes with the fewest intervening links, i.e., assuming that the relative values of variable sensitivity can be told strictly from model depth.

Why this is a problem: This misconception can incorrectly guide the expected influence of variables, and can lead to inappropriately structured models to "force" desired and anticipated sensitivity structures.

Practical solution: Realize that variable sensitivity is a function of both the model structure – namely the number of links between

Table 2
Potential objectives for a Bayesian network model.

Objective	Description
Prediction	Determine possible future outcomes based on initial conditions
Forecast	Determine the most likely future outcome based on initial conditions
Projection	Determine possible future outcomes based on changing future conditions
Scenario planning	Peg the corners of the implications of hypothetical situations
Represent knowledge	Synthesize what we think we know
Identify uncertainties and key data gaps	Identify factors or interactions with the greatest influence on outcomes; sensitivity analysis
Diagnosis	Determine potential causes of a known or specified condition or outcome
Mitigation	Identify alternative conditions that could lead to a desired outcome
Group involvement	Aid individual or collaborative decision-making; engage stakeholders

a node and the outcome – *and* the specific values of the probability tables of the variable of interest, of any intervening variables, and of the final outcome variable. A response variable could very well be more sensitive to a variable several links away than to an immediately-linked variable. When structuring and parameterizing a model from expert knowledge, use sensitivity analysis during the course of model construction to tune desired model performance by ensuring that the sensitivity of model outcomes conforms to the experts' expectations.

Issue: Conflating correlation with causation.

Why this is a problem: This is a potential, and classic, problem in models intended to be used for diagnosis or prediction, because correlates can misrepresent the true mechanics and dynamics of a system and result in incorrect model outcomes. True causation can also be misinterpreted because of lack of attention to latent, lurking, and confounding variables.

Practical solution: First clearly state the objective for the model to begin with (Table 2), as this will determine how correlates should be interpreted and included in the model. Then re-examine the network influence structure of the model as a causal web, if the model is indeed intended to represent causality. Seek peer review of the network influence structure as a causal framework and amend it accordingly. Address and represent the potential roles of latent, lurking, and confounding variables as discussed in the previous section.

Issue: Inappropriately conflating proportion with probability.

Why this is a problem: This is not a problem in cases where the counts of events or conditions in various states can be interpreted either in a sampling design or used to determine the fraction of events, conditions, areas, time, etc., as prior or conditional states. But it is a problem – it is inappropriate – to use counts of such events or conditions as proportions when: (1) the full range of possible events or conditions are not included; (2) when such counts are so few as to introduce unknown bias or inaccuracy when converting to proportions; (3) when some dynamic causality of the events or conditions are not accounted for that could belie using observed proportions as valid representations of expected values; or (4) when a BN model is intended to represent logical, syntactical, or strictly managerial (decision) systems, in which case proportions of cases having various decision outcomes may not be appropriate indicators of correct decision pathways; that is, the proportions of various decision cases may not reflect the underlying frequency of circumstances, so may not be appropriate to use as probabilities in the model.

Practical solution: Clarify the objective for the modeling exercise, then clearly define the intent and purpose of each linkage in the model. Then, identify those linkages intended to represent causal probabilities of spatial, temporal, or directly conditional events or characteristics; these may be candidates for using frequencies and proportions to inform conditional probabilities, if such data or expert knowledge are available.

Issue: Inappropriately conflating expectation with probability.

Why this is a problem: Just because an outcome can be "expected" to occur for whatever observer experience or reasons, is not necessarily indicative of its true probability of occurrence given prior conditions. This can lead to severely biased CPT entries and ill-functioning models. This can occur because people are more apt to focus on more extreme events and outcomes even if they are far less frequent than more mundane outcomes that occur more commonly (Burgman, 2016).

Practical solution: Identify when expected outcomes are based on observer biases, lack of sufficient case knowledge, and vagaries of observational conditions. Also beware of other potential sources of bias in expectations, such as with the "anchoring" of judgments to inappropriate comparisons (Hysenbelli et al., 2013), "bandwagoning" of judgments to fit others' (e.g., Sun and Müller, 2013; Friedkin et al., 2016), and other sources of judgment bias. Methods for reducing or avoiding biases in expectations include identifying potential sources, structuring well-focused knowledge elicitation queries objectively, using a Delphi elicitation technique with expert panels to explicitly reveal the basis of judgments, and other approaches (Ayyub 2001; Burgman 2016).

Issue: Assuming that the terms expert opinion, judgment, experience, and knowledge are synonymous.

Why this is a problem: Some model users (e.g., managers) and detractors may view "opinion" and "judgment" as arbitrary and perhaps even capricious belief without evidence. Oxford English Dictionary defines "opinion" as "A view or judgment formed about something, not necessarily based on fact or knowledg"[1]; and "judgment" in part as "An opinion or conclusion".[2] Generally, this is not the sort of information upon which BN models are to be based if they are to achieve scientific credibility, robustness, and acceptability.

Practical solution: Do not develop models based solely on unfounded or poorly-founded opinion or judgment, but strive to use rigorous procedures of elicitation of expert experience and knowledge (Martin et al., 2012; Krueger et al., 2012; O'Hagan et al., 2006). Define what would constitute an "expert" for any modeling project (Caley et al., 2014). Building BN models using expert-guided structuring and parameterization of probability values should be based on highest standards of expert knowledge elicitation, resulting in models that are peer reviewed and replicable, and that garner high scientific credibility. If so, do not then use "expert opinion" or "expert judgment" to describe the basis for the model; instead, use

[1] http://www.oxforddictionaries.com/us/definition/american_english/opinion.
[2] https://en.oxforddictionaries.com/definition/us/judgment.

"expert experience" or "expert knowledge." Further, when based on expert experience or expert knowledge, describe such models as 'Bayesian networks' and not "Bayesian belief networks," although my own early writings have used the latter.

4. Conclusion

This is by no means an exhaustive compendium of potential BN modeling problems, nor the only set of potential solutions. For instance, I have not addressed here problems associated with hierarchical BN modeling, time-expansion dynamic network structures, conflating variable sensitivity with variable influence, overlooking how Type I and Type II errors differ and combine into overall model classification error, interpretation of causal and correlational relationships in naive Bayes model structures, and additional problems of BN model overfitting. I invite BN modeling researchers to innovate their own tractable solutions and share ideas with others (e.g., via on-line forums such as the Australasian Bayesian Network Modelling Society, abnms.org/forum). Together, we can all help the field of BN modeling to achieve aspirations of avoiding common quandaries and problems, to become more credible, and to find more solutions of general applicability.

Acknowledgments

This paper is based on my keynote address to the Australasian Bayesian Network Modelling Society Seventh Annual Conference, 23 November 2015, at Monash University, Melbourne, Australia. My thanks to the Society administrative board, including Kevin Korb and Steven Mascaro for the invitation and opportunity, and to Steven Pawson and Scion, New Zealand, for securing financial support for my participation. Many ideas presented here resulted from discussions and collaborations with many BN model researchers and users over the years, to whom I am most grateful. I thank Trent Penman, Owen Woodberry, and two anonymous reviewers for helpful comments on the manuscript. Use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Amstrup, S.C., Marcot, B.G., Douglas, D.C., 2008. A Bayesian network modeling approach to forecasting the 21st century worldwide status of polar bears. In: DeWeaver, E.T., Bitz, C.M., Tremblay, L.-B. (Eds.), Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications. Geophysical Monograph 180. American Geophysical Union, Washington, D.C, pp. 213–268.
Arhonditsis, G.B., Stow, C.A., Steinberg, L.J., Kenney, M.A., Lathrop, R.C., McBride, S.J., Reckhow, K.H., 2006. Exploring ecological patterns with structural equation modeling and Bayesian analysis. Ecol. Modell. 192, 385–409.
Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surveys 4, 40–79.
Ayyub, B.M., 2001. Elicitation of Expert Opinions for Uncertainty and Risks. CRC Press, Boca Raton, Florida (328 pp.).
Bromley, J., 2005. Guidelines for the Use of Bayesian Networks as a Participatory Tool for Water Resource Management. MERIT – Management of the Environment and Resources Using Integrated Techniques. Centre for Ecology and Hydrology, Wallingford, U.K (118 pp. + appendix).
Burgman, M.A., 2016. Trusting Judgements: How to Get the Best Out of Experts. Cambridge University Press, Cambridge, U.K (203 pp.).
Caley, M.J., O'Leary, R.A., Fisher, R., Low-Choy, S., Johnson, S., Mengersen, K., 2014. What is an expert? A systems perspective on expertise. Ecol. Evol. 4 (3), 231–242.
Capmourteres, V., Anand, M., 2016. Assessing ecological integrity: a multi-scale structural and functional approach using Structural Equation Modeling. Ecol. Indic. 71, 258–269.
Cawley, G.C., Talbot, N.L.C., 2007. Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters. J. Mach. Learn. Res. 8, 841–861.
Chen, S.H., Pollino, C.A., 2012. Good practice in Bayesian network modelling. Environ. Modell. Software 37, 134–145.
Condamin, L., Louisot, J.-P., Naim, P., 2006. Risk Quantification: Management, Diagnosis and Hedging. John Wiley & Sons, West Sussex, England (271 pp.).
Conklin, J., 2005. Dialogue Mapping: Building Shared Understanding of Wicked Problems. John Wiley & Sons, New York (266 pp.).
Conrady, S., Jouffe, L., 2015. Bayesian Networks & BayesiaLab. Bayesia USA, Franklin, Tennessee (385 pp.).
Darwiche, A., 2009. Modeling and Reasoning with Bayesian Networks. Cambridge University Press, New York (560 pp.).
Das, B., 2000. Representing Uncertainty Using Bayesian Networks. Department of Defence, Defence Science & Technology Organization, Salisbury, Australia (58 pp.).
Delbecq, A.L., van de Ven, A.H., Gustafson, D.H., 1986. Group Techniques for Program Planning: a Guide to Nominal Group and Delphi Processes. Green Briar Press, Middleton, Wisconsin.
Dlamini, W.M., 2011. A data mining approach to predictive vegetation mapping using probabilistic graphical models. Ecol. Inf. 6, 111–124.
Fenton, N., Neil, M., 2012. Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, Boca Raton, FL (524 pp.).
Forio, M.A.E., Landuyt, D., Bennetsen, E., Lock, K., Nguyen, T.H.T., Ambarita, M.N.D., Musonge, P.L.S., Boets, P., Everaert, G., Dominguez-Granda, L., Goethals, P.L.M., 2015. Bayesian belief network models to analyse and predict ecological water quality in rivers. Ecol. Modell. 312, 222–238.
Friedkin, N.E., Proskurnikov, A.V., Tempo, R., Parsegov, S.E., 2016. Network science on belief system dynamics under logic constraints. Science 354 (6310), 321–326.
Grace, J.B., Anderson, T.M., Olff, H., Scheiner, S.M., 2010. On the specification of structural equation models for ecological systems. Ecol. Monogr. 80 (1), 67–87.
Gregory, R., Long, G., Colligan, M., Geiger, J.G., Laser, M., 2012. When experts disagree (and better science won't help much): Using structured deliberations to support endangered species recovery planning. J. Environ. Manage. 105, 30–43.
Guillemette, M., Larsen, J.K., 2002. Postdevelopment experiments to detect anthropogenic disturbances: the case of sea ducks and wind parks. Ecol. Appl. 12 (3), 868–877.
Han, M., Wang, Z., Liu, Y., 2014. Selecting one dependency estimators in Bayesian network using different MDL scores and overfitting criterion. J. Inf. Sci. Eng. 30, 371–385.
Hobbs, N.T., Hooten, M.B., 2015. Bayesian Models: A Statistical Primer for Ecologists. Princeton University Press, Princeton and Oxford (320 pp).
Holmes, D.E., Jain, L.C. (Eds.), 2010. Innovations in Bayesian Networks. Theory and Applications. Springer, New York (322 pp.).
Hoshino, E., van Putten, I., Girsang, W., Resosudarmo, B.P., Yamazaki, S., 2016. A Bayesian belief network model for community-based coastal resource management in the Kei Islands, Indonesia. Ecol. Soc. 21 (2) (art. no. 16).
Hsu, C.-C., Sandford, B.A., 2007. The Delphi technique: making sense of consensus. Pract. Assess. Res. Eval. 12 (10) http://pareonline.net/getvn.asp?v=12&n=10.
Hysenbelli, D., Rubaltelli, E., Rumiati, R., 2013. Others' opinions count, but not all of them: anchoring to ingroup versus outgroup members' behavior in charitable giving. Judgment Decis. Making 8 (6), 678–690.
Jakulin, A., Rish, I., 2006. Bayesian learning of Markov network structure. Lect. Notes Comput. Sci. 4212, 198–209.
Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs, 2nd edition. Springer (448 pp.).
Jiang, L., Cai, Z., Wang, D., Zhang, H., 2012. Improving tree augmented naive Bayes for class probability estimation. Knowledge-Based Syst. 26, 239–245.
King, R.S., Baker, M.E., Whigham, D.F., Weller, D.E., Jordan, T.E., Kazyak, P.F., Hurd, M.K., 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecol. Appl. 15 (1), 137–153.
Kjaerulff, U.B., Madsen, A.L., 2007. Bayesian Networks and Influence Diagrams: a Guide to Construction and Analysis. Springer (318 pp.).
Korb, K.B., Nicholson, A.E., 2010a. Bayesian Artificial Intelligence, second edition. Chapman & Hall/CRC Press, Boca Raton, Florida (491 pp.).
Korb, K.B., Nicholson, A.E., 2010b. Knowledge engineering with Bayesian networks. In: Korb, K.B., Nicholson, A.E. (Eds.), Bayesian Artificial Intelligence. , second edition. Chapman & Hall/CRC Press, Boca Raton, Florida, pp. 297–359.
Korb, K.B., Twardy, C.R., Handfield, T., Oppy, G., 2005. Causal Reasoning with Causal Models Clayton School of Information Technology. Monash University, Clayton, Victoria, Australia (45 pp.).
Koski, T., Noble, J., 2011. Bayesian Networks: An Introduction. Wiley (366 pp.).
Krueger, T., Page, T., Smith, L., Vionov, A., 2012. A guide to expert opinion in environmental modelling and management. Environ. Modell. Software 36, 1–3.
Kuschner, K.W., Malyarenko, D.I., Cooke, W.E., Cazares, L.H., Semmes, O.J., Tracy, E.R., 2010. A Bayesian network approach to feature selection in mass spectrometry data. BMC Bioinf. 11, 177, http://dx.doi.org/10.1186/1471-2105-11-177.
Landuyt, D., Broekx, S., Goethals, P.L.M., 2016. Bayesian belief networks to analyse trade-offs among ecosystem services at the regional scale. Ecol. Indic. 71, 327–335.
Marcot, B.G., Holthausen, R.S., Raphael, M.G., Rowland, M.M., Wisdom, M.J., 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement? For. Ecol. Manage. 153 (1–3), 29–42.
Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Can. J. For. Res. 36, 3063–3074.
Marcot, B.G., 1990. Representing uncertainty in expert systems: I. Use of likelihood estimates in knowledge base rules. Nat. Resour. Comp. Newslett. 5 (9), 10–14.

Marcot, B.G., 1991. Representing uncertainty in expert systems: II. Assumptions and limitations of the likelihood approach. Nat. Res. Comp. Newslett. 6 (1), 8–10.

Marcot, B.G., 2007. In: Naim, P., Wuillemin, P.-H., Leray, P., Pourret, O., Becker, A. (Eds.), Étude De Cas n°5: Gestion De Ressources Naturelles Et Analyses De Risques (Natural Resource Assessment and Risk Management). Réseaux Bayésiens (Bayesian networks; in French), Eyrolles, Paris, France, pp. 293–315.

Marcot, B.G., 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecol. Modell. 230, 50–62.

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M., Kengersen, K., 2012. Eliciting expert knowledge in conservation science. Conserv. Biol. 26 (1), 29–38.

Moe, S.J., Haande, S., Couture, R.-M., 2016. Climate change, cyanobacteria blooms and ecological status of lakes: a Bayesian network approach. Ecol. Modell. 337, 330–347.

Myllymäki, P., Silander, T., Tirri, H., Uronen, P., 2002. B-Course: A web-based tool for Bayesian and causal data analysis. Int. J. Artif. Intell. Tools 11, 369–387.

Neopolitan, R.E., 2003. Learning Bayesian Networks. Prentice Hall, New York (674 pp.).

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain Judgements: Eliciting Experts' Probabilities. Wiley, New York (338 pp.).

Oniśko, A., Druzdzel, M.J., Wasyluk, H., 2001. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. Int. J. Approximate Reasoning 27 (2), 165–182.

Orsi, F., Geneletti, D., Newton, A.C., 2011. Towards a common set of criteria and indicators to identify forest restoration priorities: an expert panel-based approach. Ecol. Indic. 11 (2), 337–347.

Park, S., Budescu, D.V., 2015. Aggregating multiple probability intervals to improve calibration. Judgment Decis. Making 10 (2), 130–143.

Pearl, J., 2000. Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge (384 pp.).

Pitchforth, J., Mengersen, K., 2013. A proposed validation framework for expert elicited Bayesian Networks. Expert Syst. Appl. 40 (1), 162–167.

Pollino, C.A., Henderson, C., 2010. Bayesian Networks: A Guide for Their Application in Natural Resource Management and Policy. Landscape Logic Technical Report No. 14. Department of the Environment, Water, Heritage and the Arts, Australian Government (48 pp.).

Pollino, C.A., Woodberry, O., Nicholson, A., Korb, K., Hart, B.T., 2007. Parameterization and evaluation of a Bayesian network for use in ecological risk assessment. Environ. Modell. Software 22 (8), 1140–1152.

Pourret, O., Naïm, P., Marcot, B.G. (Eds.), 2008. Bayesian Belief Networks: a Practical Guide to Applications. Wiley, West Sussex, England (432 pp.).

Qin, B., Xia, Y., Wang, S., Du, X., 2011. A novel Bayesian classification for uncertain data. Knowledge-Based Syst. 24 (8), 1151–1158.

Richards, R., Meynecke, J.-O., Sahin, O., Tiller, R., Liu, Y., 2014. Ocean acidification and fisheries – a Bayesian network approach to assessing a wicked problem. In: Ames, D.P., Quinn, N.W.T., Rizzoli, A.E. (Eds.), 7th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs), San Diego, California (1998–2005 pp.).

Rittel, H.W.J., Webber, M.M., 1973. Dilemmas in a general theory of planning. Policy Sci. 4, 155–173.

Rothman, K.J., Greenland, S., 1998. Modern Epidemiology, second edition. Lippincott-Raven, Philadelphia, PA.

Sethi, S.A., Dalton, M., 2012. Risk measures for natural resource management description, simulation testing, and R code with fisheries examples. J. Fish Wildlife Manage. 3 (1), 150–157.

Sloman, S., 2009. Causal Models: How People Think About the World and Its Alternatives. Oxford University Press (224 pp).

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L., Cowell, R.G., 1993. Bayesian analysis in expert systems. Stat. Sci. 8 (3), 219–283.

Stevens, R.H., Thadani, V., 2006. A Bayesian network approach for modeling the influence of contextual variables on scientific problem solving. Lect. Notes Comput. Sci. 4053, 71–84.

Sun, Z., Müller, D., 2013. A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. Environ. Modell. Software 45, 15–28.

Uusitalo, L., Kuikka, S., Romakkaniemi, A., 2005. Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. ICES J. Mar. Sci. 62, 708–722.

Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling? Ecol. Modell. 203 (3–4), 312–318.

Van Allen, T., Singh, A., Greiner, R., Hooper, P., 2008. Quantifying the uncertainty of a belief net response: bayesian error-bars for belief net inference? Artif. Intell. 172 (4–5), 483–513.

Wilhere, G.F., 2012. Using Bayesian networks to incorporate uncertainty in Habitat Suitability Index models. J. Wildl. Manage. 76 (6), 1298–1309.

Woodberry, O., Mascaro, S., 2014. Programming Bayesian Network Solutions with Netica. Bayesian Intelligence (204 pp.).

van Bueren, E.M., Klijn, E.-H., Koppenjan, J.F.M., 2003. Dealing with wicked problems in networks: analyzing an environmental debate from a network perspective. J. Public Adm. Res. Theory 13 (2), 193–212.