Advances in Bayesian network modelling: Integration of modelling technologies

Bruce G. Marcot, Trent D. Penman

USDA Forest Service, Portland, OR, USA
School of Ecosystem and Forest Sciences, University of Melbourne, Victoria, Australia

ARTICLE INFO

Keywords:
Bayesian networks
Decision models
Model integration
Machine learning
Model validation

ABSTRACT

Bayesian network (BN) modeling is a rapidly advancing field. Here we explore new methods by which BN model development and application are being joined with other tools and model frameworks. Advances include improving areas of Bayesian classifiers and machine-learning algorithms for model structuring and parameterization, and development of time-dynamic models. Increasingly, BN models are being integrated with: management decision networks; structural equation modeling of causal networks; Bayesian neural networks; combined discrete and continuous variables; object-oriented and agent-based models; state-and-transition models; geographic information systems; quantum probability; and other fields. Integrated BNs (IBNs) are becoming useful tools in risk analysis, risk management, and decision science for resource planning and environmental management. In the near future, IBNs may become self-structuring, self-learning systems fed by real-time monitoring data. Such advances may make model validation difficult, and may question model credibility, particularly if based on uncertain sources of knowledge systems and big data.

1. Introduction

Bayesian networks (BNs) are directed acyclic graphs that link variables by conditional probabilities, where model outputs are probabilities calculated using Bayes’ Theorem (Fenton and Neil, 2012; Koski and Noble, 2011). BN modeling is useful for data mining, determining and explicitly displaying the relationship among variables, representing expert knowledge and combining expert knowledge and empirical data, and identifying key uncertainties (Cheon et al., 2009; Hanea et al., 2010; Landuyt et al., 2013). Outputs are typically expressed as probabilities of various states, which lends well to decision-science approaches to risk analysis and risk management (Aalders, 2008; Farmani et al., 2012).

General network structure of BN models is highly flexible, leading many researchers to find new areas of application. As examples, BNs have become popular in environmental management for projecting potential impacts of proposed projects (Krüger and Lakes, 2015), forecasting impacts of environmental disturbances such as fire (Dlamini, 2010) and climate change (Sperotto et al., 2017) and providing a basis for making environmental management decisions (Barton et al., 2012). Many examples are available of the use of BNs in a wide variety of other environmental and resource management contexts, such as management of groundwater (Giordano et al., 2013), recreation impacts (Fortin et al., 2016), and green energy production (Carta et al., 2011). If the BN contains no random variable, then the outcome generated is fixed, i.e., deterministic, for a given set of priors, else the outcome is stochastic. BNs can be made stochastic by introducing random deviates as part of formulae within nodes. Variables also can be described with formulae combining values of parent nodes, such as used by Steventon et al. (2006) in assessing viability risk of a rare seabird. Further, variables can be denoted with continuous ranges, rather than discrete state conditions, such as used by Hradsky et al. (2017) to determine impacts of fire and other stressors on the distribution of terrestrial wildlife.

BNs are a highly useful tool for depicting and modeling current knowledge, such as with initial representation of a system or problem to gain a better understanding and perspective on uncertainties and complexities so as to help advise managers and decision-makers. BNs provide a robust statistical framework when little data are available. An example is in the context of environmental modeling with time-critical situations with scant available data, such as active monitoring of key energy infrastructures (Guerriero et al., 2016) and surveillance of endangered species (Koen et al., 2017).

Increasingly, BNs are being integrated with other modeling constructs and tools, such as geographic information systems (GIS) and remote sensing databases. In this paper, we explore these new avenues
of how BN model development and application are being joined with many other tools and model frameworks for a variety of environmental assessment and management objectives. We briefly review the current state and recent advances of BN modeling, and then provide examples of an emerging new era of integrating BN models with other frameworks and tools. Lastly, we present a vision of next advances to come, concluding with a perspective on ensuring scientific and decision-making credibility, with cautions on accelerated model advancements.

2. The cutting edge of BN modeling

The field of BN modeling is advancing swiftly with the number of journal articles using BNs continuing to rise, including a recent era of exponential growth (Marcot, 2017). Recent developments in BN modeling are reviewed in the following sections.

2.1. Recent advances in BN model structure and applications

A number of recent advances in BN model structure and application have followed a diverse track of topics, generally related to identifying and exploring system dynamics and aiding decision management (Fig. 1).

Classification. One area of resurgence is in new approaches to the classification problem, viz., Bayesian classifiers and machine learning. BN classifiers include a wide variety of algorithms starting with naive Bayes and variants thereof (Bielza and Larrañaga, 2014). Related to this are algorithms for Bayesian learning of probability structures from empirical data (e.g., Tsamardinos et al., 2006; Do and Batzoglou, 2008).

Latent variables. A common problem in ecological or environmental modeling is the influence of latent variables, which are effects inferred from the relation among observed variables but which are not directly observed (Marcot, 2017). Machine-learning algorithms used in parameterizing the probability values in BN models, such as the expectation maximization algorithm (Do and Batzoglou, 2008), can, to some degree, account for the influence of latent variables and missing data (Lauritzen, 1995).

A related problem is how to validate BN models developed entirely from expert elicitation with no case-file data by which to structure or parameterize the model. Such models portray logical or causal relations among variables as inferred by expert knowledge, but these relations often are influenced by unobserved, latent variables (de Waal et al., 2016). Pitchforth and Mengersen (2013) proposed methods for evaluating confidence in the validity of such models even in the dearth of empirical data and presence of latent variables, thus providing a validation framework for expert-elicited BNs. de Waal et al. (2016) suggested several approaches to handling latent variables in BNs, including explaining uncertainties associated with latent variables, parameterizing the probability values of BNs so as to directly address the roles of latent variables, and addressing uncertainty in model validation.

Depicting model confidence. Although BN models explicitly incorporate uncertainty, there is no common method for depicting and quantifying the degree of confidence in the underlying probability values of the model and in the resulting posterior probability calculations. That is, uncertainty measures can be inferred from the probability distributions of states calculated in the model, but these are not necessarily the same as the degree of confidence (lever of certainty) of that probability distribution. Pitchforth and Mengersen (2013) characterized confidence in BN model behavior as consisting of three components of structure confidence, discretization confidence, and parameterization confidence. For use in BN model validation, they further added 7 common dimensions of validity as used in psychometry: nomological, face, content, concurrent, predictive, convergent, and discriminant validity, which, collectively, pertain to the degree of concordance within accepted norms and credibility within a particular discipline.

A more quantitative method of depicting BN model confidence, as developed by Van Allen et al. (2008), entails estimating error bars around posterior probability calculations from BNs, which then depict the degree of uncertainty (or confidence) in model outcomes. Error bars for BNs are termed credible intervals, which provide the range of model outcomes within a specified probability level (Marcot, 2012), but which are not often reported in BN modeling projects. Such error bars should not be confused with frequentist confidence intervals. Hamilton et al. (2015) used credible intervals to measure the strength of the relationship between suitability of habitat of a crayfish and environmental predictor variables.

Links to GIS. Probability outcomes from BN models for evaluating local conditions have been used as input to GIS systems to create maps depicting habitat quality of wildlife species (Raphael et al., 2001; Havron et al., 2017), Kininmonth et al. (2014) presented a model which combined spatial datasets, spatial models, and expert opinion in an integrated BN-GIS structure for evaluating boating damage to the Great Barrier Reef of eastern Australia. Diamini (2010) developed a BN-GIS model that uses geographically-referenced remote sensing MODIS data to analyze wildfire in Swaziland. BN programs that integrate or intersect with GIS include GeoNetica® (Norsys Inc.), HUGIN® (HUGIN Expert A/S), and Ecosystem Management Decision Support (EMDS, Reynolds et al., 2014) that integrates the GeNiE BN modeling platform with two GIS components of ArcMap® (Esri) and open-source QGIS. Gonzalez-Redin et al. (2016) BNs linked to GIS to map trade-offs of ecosystem services in the French Alps to inform planning decisions. Several projects have explicitly integrated GIS and BN modeling frameworks, such as the QGIS plug-in for BNs developed by Landuyt et al. (2015) and the integration of the GeNe BN modeling framework into the ArcGis®-based Ecosystem Management Decision Support system (emds.mountainviewgroup.com).

Dynamic Bayesian networks. Other recent variations on the traditional BN modeling theme include dynamic Bayesian networks (DBNs) that model a time series of conditions and contingencies, such as with oscillating predator-prey dynamics (Fig. 2). DBNs typically contain feedback loops which are not allowed in the directed acyclic graph structure of BNs, but can be modeled when BNs are time-expanded so that the entire BN structure is replicated for different time periods so that the links become acyclic. In some cases, DBNs have been made spatially explicit by integrating with geographic information systems (GIS; e.g., Chee et al., 2016). A variant of DBNs are those operating in real time in response to discrete or continuous inputs, such as for predicting highway crashes (Hossain and Muromachi, 2012) and in analyzing gene networks (Kim et al., 2003). New approaches to structuring
DBNs combine methods from static and dynamic networks (Vlasselaer et al., 2016). Uusitalo et al. (2018) used DBNs with hidden variables to model major structural changes of a Baltic Sea food web, and Orphanou et al. (2014) used “temporal Bayesian networks” (TBNs, a synonym for DBN) to evaluate temporal relationships in clinical data for medical diagnosis and prognosis. Their hidden variables represented unobserved processes contributing to the changes and resulted in DBN models that reflected known dynamics of the food web system.

Bayesian decision networks. Bayesian decision networks (BDNs) extend BN models by explicitly including decision and utility nodes (e.g., Barton et al., 2008). Decision nodes are deterministic nodes that depict unique management decisions, and utility nodes are continuous nodes that estimate a cost or benefit of a given outcome resulting from a decision. BDNs use utility nodes to calculate overall expected values of all costs or benefits of alternative management decisions, given the probability structure of the model, and can be highly useful in risk analysis and risk management arenas. For example, Loyd and DeVore (2010) developed a BDN to advise on alternatives for management of feral cats in the United States. Catenacci and Guipponi (2013) used a BDN as a basis for adaptation planning to sea-level rise. A further variation of BDNs is with dynamic decision networks (DDNs) that essentially merge decision networks with time-expanded dynamic networks. DDNs were developed by Murray et al. (2004) to guide selection of teaching tutorials, and by Penman et al. (2015a) to advise on reducing risk of loss of homes to wildfire. BNs have also been used to assess value of information to optimize resource use decisions, such as the fisheries industry (Kuikka et al., 1999).

Depicting causality in structural equation models. BN modeling has been compared to structural equation modeling (SEM) in that both can be used to depict causal networks and influences and can analyze degree of causality (Pearl, 1998, 2000). The two approaches differ in that SEM is a general suite of statistical tools usually using frequentist, multivariate approaches (although some SEM approaches also support Bayesian estimation), whereas BNs use conditional probabilities and Bayes’ theorem. A main difference is that SEMs are purely statistical tools developed, for example, to test hypotheses or to test whether an assumed causal relation in the graph is significant, whereas BNS use probabilistic models (trainable by data) mainly for investigating the consequences of conditions or events on outcomes, or deducing causal conditions resulting in an outcome.

More recently, Li et al. (2018) compared and combined BN and SEM modeling to evaluate the interactive influence of land use and climate
change on stream macroinvertebrates. They first used SEM to develop a conceptual influence diagram of causal effects, that is, the network structure of variables and their linkages, and then they built prediction and diagnostic BNs from the same conceptual model. Using SEM helped identify and justify the links used in the BN models. Their results suggested that modeling all causal factors together in the SEM conceptual model and in the subsequent BN model provided a more robust understanding of how positive effects from climate change could mollify negative influences from land use.

Neural networks. A somewhat different variant of BN modeling appears in Bayesian neural nets, that is, using Bayesian learning to determine neural network node weights. Neural networks are typically trained using a variety of approaches including variants of gradient descent and least-squares methods to minimize loss functions of variables singly or in conjunction. Using BNs can bring greater efficiency in adjusting node weight parameters based on prior knowledge. Bayesian neural networks have been used to forecast energy load requirements (Lauret et al., 2008), solar irradiation (Yacef et al., 2012), stock market performance (Ticknor, 2013), and internet traffic loads (Auld et al., 2007).

Continuous Bayesian networks. Another area of recent interest and progress is in developing continuous BN models where quantitative variables are not discretized into exclusive state ranges but instead are represented by continuous values such as equations or statistical distributions. Continuous BNs can be constructed using programming tools such as UNINET (Cooke et al., 2007; Delgado-Hernández et al., 2012), WinBUGS (Kery, 2010), AgenaRisk® (Neil et al., 2007), Hugin® (Madsen et al., 2003), and GeNte (Druzdzel, 1999). Other BN modeling and graphical modeling software also can deal with continuous nodes if the multivariate normal assumption is made (A. Hanea, pers. comm.)

Hybrid Bayesian networks. Some researchers have developed BN models with both discrete and continuous variables where the latter are not discretized (Aguilera et al., 2016; Castillo et al., 1998; Driver and Morrell, 1995). These types of models are referred to as hybrid BNs (HBNs; Hanea et al., 2006). A special case of HBNs, called non-parametric BNs (NPBNs), was reviewed by Hanea et al. (2015). NPBNs were initially devised for continuous-only BNs but are used in situations of HBns as well. Hanea et al. (2010) developed a NPBN methodology with data mining to develop prediction models. Hradsky et al. (2017) used NPBNs to model a presence-absence continuous response of wildlife to fire age classes and terrestrial vegetation classes depicted with discrete variables.

Object-oriented Bayesian networks. A further area of recent exploration is with object-oriented Bayesian networks (OOBNs) and dynamic object-oriented Bayesian networks (DOOBNs) (Bangsø et al., 2004; Benjamin-Fink and Reilly, 2017). For example, OOBNs and DOOBNs have been used to model health impacts of cyanobacteria blooms (Johnson et al., 2010), viability of populations of cheetahs (Acinonyx jubatus) in Namibia (Johnson et al., 2013), and issues of water resource management (Phan et al., 2016). The tools Hugin® and AgenaRisk® provide for true OOBN modeling, and GeNte® also can be used as such although it is not a true OOBN framework.

Agent-based modeling. Related to OOBNs is the merging of agent-based models and BNs. Agent-based models (An, 2012) are simulations of the dynamics, such as movement patterns, of individual objects. Nielsen and Parsons (2007) developed a model of consensus-building where individual agents were represented by BNs that expressed a range of possible agreements. Sun and Müller (2013) presented a BN agent model to explore the economics of ecosystem services and land-use decision-making.

State-and-transition modeling. Another area of integration is with state-and-transition models (STMs) that are used to project future proportions or amounts of conditions, such as landscape vegetation conditions and species responses, under known or hypothesized rates of change (Mason et al., 2017). STMs project future conditions, such as area covered in vegetation type categories, by multiplying a matrix of current area in each category by a matrix of probabilities depicting transitions to the same or other categories (e.g., Jorgenson et al., 2015). In a hybrid STM-BN model, transition probabilities are estimated from calculations in the BN network that account for environmental influences on each vegetation type category. Bashari et al. (2009) developed an integrated STM-BN model, as expanded upon by Nicholson and Flores (2011), to inform management decisions in rangelands of Queensland, Australia. Chee et al. (2016) integrated STMs and BNs in a geographic information system (GIS), with object-oriented concepts, to model spatial and temporal changes in an Australian woodland and a wetland in Florida.

Quantum Bayesian networks. BNs are being increasingly used in the area of quantum information theory as quantum Bayesian networks (QBNs; Tucci, 1995). QBNs are constructed to represent outcomes that deviate from, and are paradoxical to, classical probability calculations. Examples include when outcomes are dependent on the sequence of inputs (parents and parent nodes in a BN); when human decision-making deviates from dominant probability outcomes in a BDn; when a system can result in >1 dominant probability outcome state; and other situations. Such outcomes could be modeled in traditional BNs by including latent variables but such models quickly become overly complex and serve only to describe specific conditions and outcomes, not to serve as predictive and explanatory models.

Generally, in QBNs, classical conditional probability tables using Bayes calculus are replaced by quantum probability amplitudes (a complex number function that describes the behavior of a system). Moreira and Wichert (2018) developed a decision-based QBN with quantum probability amplitudes to demonstrate how prediction of some aspects of human decision-making is more efficient than with a traditional BN with latent variables. Similarly, Trueblood et al. (2016) used a QBN approach to model how human judgment can deviate from classical probability in the face of high uncertainty and imperfect information about causality of a system. Busemeyer and Trueblood (2009) explored the use of QBNs in quantum theory to model how different sequences of measurements can affect the probabilities of system outcomes. Leifer and Spekkens (2013) developed a QBN framework for depicting how quantum conditional states can result from the influence of two systems at one time or from one system at two times. Other formulations and applications of QBNs are found in the literature, although at present there does not seem to be any generally available software by which QBNs can be constructed.

Power PC theory and causal BNs. Power PC (“probabilistic contrast”) theory states that a system outcome is the sum effect of the relative power of observed and unobserved causal relations which can be depicted and partitioned mathematically (Cheng, 1997; Norick and Cheng, 2004). In applying power PC theory, Lu et al. (2008) demonstrated how the relative influence of different, independent causes can be determined empirically and can be represented in Bayesian causal networks. Here, we have covered a range of significant recent advances in application of BNs. Still other variations and new approaches to BN modeling continue to appear in the literature.

2.2. Beyond the network: a new era of integration

BN models of various forms have been increasingly used in a variety of applications. They are also being specifically integrated with other modeling constructs, which we refer to here as integrated Bayesian networks (IBNs; Johnson et al., 2010). IBNs can be defined as BN structures that are explicitly embedded within the framework of other modeling constructs, instead of just being applied to some area of inquiry as reviewed in the previous section. Examples of IBNs include assimilating BNs in structured decision-making frameworks, agent-
based models, and state and transition models (Table 1). IBNs are crafted generally to apply the probabilistic basis of BNs to new areas of application and research such as dynamic and stochastic simulation modeling.

IBNs are also becoming useful tools in risk analysis, risk management, and decision science, such as in environmental resource planning (Johnson and Mengersen, 2012; Fraser et al., 2017) and for depicting how deep uncertainty affects policies decisions (Aven, 2013; Cox, 2012). Janssens et al. (2006) developed an IBN that combined BNs and decision trees for developing decision rules in transportation management. QGeNLe Modeler® (BayesFusion, LLC) has a graphical user interface that provides for rapid prototyping of decision models in a BN environment. Some BN modeling platforms such as Netica® (Norsys Inc.) and Hugin® provide application programming interfaces (APIs) to facilitate integration links to other programs such as geographic information systems which can facilitate their use in integrated risk analysis and structured decision-making under uncertainty (e.g., Barton et al., 2008).

An emerging area is the development of IBNs operating from real-time monitoring data. For example, Maglogiannis et al. (2006) have proposed a patient-health risk analysis system using IBNs operating from vital sign monitoring data. Their vision is to produce a real-time system for homecare telemedicine. Penman et al. (2015b) developed a fire danger rating system that updated daily with meteorology forecasts, and as new fires appeared in the landscape the model automatically updated the risk projections. Vagnoli et al. (2017) proposed a real-time, IBN-based system to monitor the structural integrity of railway bridges in Europe. Koen et al. (2017) developed an IBN model to monitor the poaching (illegal hunting) of rhinoceroses in Kruger National Park, South Africa, which is being implemented as a real-time management tool.

3. Things to come

The field will continue to advance more rapidly into uncharted territory as IBNs become more sophisticated. Future IBN research foci will include substantial cutting-edge advancements in the many areas of integration reviewed above. Currently, IBNs are essentially static; however, we expect that in the near future, IBNs will emerge that are self-updating and self-improving, and that will learn from real-time continuous input of environmental monitoring data. Being able to dynamically update BN conditional probability values with new data (e.g. using an expectation maximization algorithm), and to continually re-calculate posterior probabilities, are quite feasible with existing software and hardware. Some current IBNs do automatic recalculation of posterior probability outcomes with new input data feeding the models, but here we are referring also to the basic structure and underlying conditional probability tables themselves being created anew and updated with machine-learning tools. The purpose of this would be to continually refine the context, accuracy, and robustness of the models, which is a fundamental precept and advantage of Bayesian statistical approaches that improve model predictions from prior data. Some aspects of this have appeared in the application of BNs in areas of finance (Giudici and Spelta, 2016; Garvey et al., 2015), and medical diagnosis and decision-support (Constantinuo et al., 2016). In the future, we envision self-creation of IBNs based on emergent information from crowdsourced data (Park and Budescu, 2015).

Also to come will be the further integration of BNs with expert system knowledge bases, particularly using fuzzy-logic or neural-net forms of knowledge-representation and machine-learning algorithms. In the past, control-rule-based expert systems used confidence factors or scoring rules (Zohar and Rozenstock, 2008), e.g., indices as provided by the expert and scaled [1,10], to rank the likelihood or the credibility of a particular inference or outcome. Today, the BN modeling shell BayesiaLab® (Bayesia S.A.S.; Conardy and Jouffe, 2015) averages conditional probability values from multiple experts and weights the values by the product of confidence in, and credibility of, the values as scored by each expert.

4. Conclusions and perspectives

What are some of the main cautions and caveats in this new era of IBNs? Here we address quandaries of ensuring validity and credibility of IBNs that are becoming increasingly complex in construction and interpretation, particularly as they are induced from automated and big data sources.

4.1. Ensuring a future of validity and credibility

It will be important to ensure the validity and credibility of increasingly-complex expert-based IBN systems (Kleemann et al., 2017), particularly as they interact with human social systems, as they guide resource management decisions, and as they operate with greater autonomy. As IBNs become more complex, it will become more problematic to create and test simple, intuitive, and understandable influence diagrams and mind maps that chart their structures, logic, and operation. A way around this could entail decomposing such complex systems into simpler component submodels, and testing and updating each submodel. BNs are generally constructed as Markov processes so that they can be dissected and reassembled without loss of information, particularly using cutpoints in the network graph (nodes in the BN whose removal would separate the graph).

Direct and indirect causes in process models will be increasingly difficult to clearly identify, particularly with models having multiple interaction terms, feedback loops, latent variables, and synergistic
functions among covariates and response variables. This also means increasing difficulty in parsing out sensitivity and influence effects to specific drivers, that is, clearly identifying key factors that most influence outcomes and for which management might have greatest uncertainty and greatest (or least) control. The result may limit their acceptance and adoption in management situations if users do not understand and trust the models and if the models have limited face validity. This can be a major issue with decision networks and use of models in real time, and with models intended to guide planning and management of resources with high opportunity costs. Support for the model, as with advances in validation methods and frameworks as discussed above, is needed. Some work advancing methods of sensitivity analysis of BN models may show promise for further development and application on increasingly complex model structures (e.g., global sensitivity analysis methods of Li and Mahadevan, 2017).

Increasing complexity of IBNs will carry increased difficulty in model calibration, testing, validation, and updating, eventually necessitating new heuristic approaches and algorithms that can wade the swamp of big data (Spiegelhalter, 2014; Deau et al., 2017; Lv et al., 2014). Lewis et al. (2018) further warned that developing wildlife sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even automated monitoring (e.g., environmental and Earth Science remote

4.2. Cautions and caveats in the new era of integration

We have tried here to chart some routes that development of IBN systems may be taking, and some cautions indicated in this fast-evolving era (Salmond et al., 2017). Soon to come will be dynamic and self-modifying and even self-creating IBNs using “big data” garnered from automated monitoring (e.g., environmental and Earth Science remote sensing data; Li et al., 2016), citizen science initiatives, and even crowdsourced information sources. We ask, to what degree should we trust such models? For, in the end, the knowledge source and expertise that may serve to generate and fuel IBNs may become a fully realized global, emergent artificial intelligence. If so, how shall the models’ veracity be determined when their source becomes non-human?

BNs developed from empirical data often can be tested using existing cross-validation and jackknifing algorithms. However, the validity of BNs developed from expert knowledge is more difficult to determine if independent data sets by which to test the models are unavailable, or if specific combinations of variables in big data sources are sparse; in such cases, peer review of the model at various stages of development is essential to establish credibility and conformity with accepted precepts. But determining the external validity of self-organizing and self-updating BNs that are developed with deep-learning algorithms in real time from big data sources, including citizen-science and crowd-sourced data (especially the multitude of Internet blog sources and news posts), without independent testing or review, may be most problematic and will require new approaches to scientifically evaluating their veracity. This will become one of the greater challenges in the fast-evolving new era of BN model integration.

Acknowledgments

Inspiration for this paper comes from a keynote address given by the senior author in 2017 at the Joint Conference of the Australasian Bayesian Network Modelling Society and the Society for Risk Analysis Australia and New Zealand, at the University of Melbourne, Australia. We thank Tom Bruce, Anca Hanea, Sandra Johnson, and two anonymous reviewers for their helpful reviews of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2018.09.016.

References

Catenacci, M., Guipponi, C., 2013. Integrated assessment of sea-level rise adaptation


