International Statistical Review

International Statistical Review (2019), 0, 0, 1–19 doi:10.1111/insr.12350

Combining Opinions for Use in Bayesian Networks: A Measurement Error Approach

A. Charisse Farr[1], Kerrie Mengersen[1], Fabrizio Ruggeri[1,2], Daniel Simpson[3], Paul Wu[1] and Prasad Yarlagadda[1]

[1]*Science and Engineering Faculty, Mathematical Sciences, Queensland University of Technology, Brisbane, Qld, Australia*
[2]*Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, Milan, Italy*
[3]*University of Toronto, Toronto, ON, Canada*
E-mail: fabrizio@mi.imati.cnr.it

Summary

Bayesian networks (BNs) are graphical probabilistic models used for reasoning under uncertainty. These models are becoming increasingly popular in a range of fields including engineering, ecology, computational biology, medical diagnosis and forensics. In most of these cases, the BNs are quantified using information from experts or from users' opinions. While this quantification is straightforward for one expert, there is still debate about how to represent opinions from multiple experts in a BN. This paper proposes the use of a measurement error model to achieve this. The proposed model addresses the issues associated with current methods of combining opinions such as the absence of a coherent probability model, the loss of the conditional independence structure of the BN and the provision of only a point estimate for the consensus. The proposed model is applied to a subnetwork (the three final nodes) of a larger BN about wayfinding in airports. It is shown that the approach performs well than do existing methods of combining opinions.

Key words: Bayesian networks; expert opinions; measurement error model; wayfinding.

1 Introduction

Bayesian networks (BNs) have become a ubiquitous statistical tool for describing complex systems. A typical BN is based on a directed acyclic graph (DAG) model in which variables are represented as nodes and probabilistically linked by a set of directed arcs. BNs are growing in popularity in engineering (Trucco *et al.*, 2008), ecology (Johnson, 2009), natural resource management (Pollino *et al.*, 2007), computational biology (Friedman *et al.*, 2000), medical diagnosis (Heckerman, 1990) and forensics (Taroni *et al.*, 2004). The application of BNs has also had an impact on their methodological aspects, raising various issues that still need to be addressed adequately. One of them is the way in which information from multiple experts, possibly with different opinions and levels of expertise, can be represented in a BN.

Common ways of addressing this issue include consensus building approaches such as the Delphi method (Dalkey & Helmer, 1963), averaging using the arithmetic mean (Beliakov *et al.*, 2007; Burgman *et al.*, 2011) or linear pooling (Cooke, 1991; French, 2011; Genest & Zidek, 1986), and Bayesian approaches (French, 1985; Lindley, 1983; West, 1988; Winkler, 1968). These are briefly discussed in Section 4.

In this paper, a BN is represented as a DAG, and the probabilities attributed to each node in the BN by each expert are assumed to be observations of the underlying *true* probabilities, subject to measurement error. The paper provides a review of measurement error models in the context of generalised linear random effects models and proposes their use in representing the systematic variation in the probabilities assigned by the experts. The novel use of measurement error models to combine expert opinions in BNs is the major contribution of the paper. Compared with linear pooling, the proposed model has the advantage of following from a coherent probability model and allowing for uncertainty, because the resulting distribution is more informative than a point estimate.

The conditional dependencies imposed by the BN are reflected in the associated precision matrices. The approach is then applied to a key subnetwork of a wayfinding BN model (WBNM) (Farr *et al.*, 2014) that was developed to investigate the factors that influence effective wayfinding in airports. Previous wayfinding research has been split into two distinct streams: research that investigated human factors such as cognition to define issues such as cognitive mapping, information processing, memory and spatial recognition (Gärling *et al.*, 1984; Kuipers, 1978; Passini, 1981; 1984; Peponis *et al.*, 1990; Timpf *et al.*, 1992), and research that used environmental factors to present mathematical measures such as the visibility index (Braaksma *et al.*, 1980; Dada & Wirasinghe, 1999; Tosic & Babic, 1984) and the interconnection density (O'Neill, 1991). The WBNM combines these previously separate sets of factors into a single model via a BN. Given their relevance, we have decided to concentrate on the final nodes of those two streams and their influence on wayfinding. Therefore, we apply the measurement error model to the subnetwork that relates the final three nodes of the WBNM, that is, human factors, environmental factors and wayfinding. The WBNM has been chosen because it is a real case study, with a significant number of experts providing their opinions, whereas the decision of concentrating on the three final nodes is justified not only by computational constraints, which would hinder how the proposed method works if a larger number of nodes were used, but also by the possibility of getting extra insights about the influence of human and environmental factors on wayfinding. The structure of the WBNM was fixed in earlier studies and not questioned by the experts: the combination of experts' opinion on structuring a BN is a very complex problem, well beyond the scope of the current paper.

The paper is structured as follows. In order to motivate the problem, the wayfinding case study and the WBNM model are described in Section 2. In Section 3, the basic properties of DAGs and BNs are presented, and existing methods for combining information in BN are described in Section 4. Measurement error and generalised linear models (GLMs) are presented in Sections 5 and 6, respectively. The proposed model is described in Section 7, illustrated for a three-node BN in Section 8 and applied to the WBNM subnetwork for the wayfinding case study in Section 9. Finally, additional comments and pointers for future research are discussed in Section 10.

2 Case Study—The Wayfinding Bayesian Network Model

Wayfinding is the 'process of finding your way to a destination in a familiar or unfamiliar setting using cues given by the environment' (Farr *et al.*, 2012). It requires the successful interplay between human and environmental factors. Previous research on wayfinding has investigated this process from one of two perspectives, namely, either human or environmental factors. Studies of human factors have investigated issues such as memory, cognitive mapping, spatial recognition and information processing (Gärling *et al.*, 1984; Kuipers, 1978; Passini, 1981). Studies of environment factors have focused on measures related to the wayfinding process. For example, the visibility index gives a measure of the ease of wayfinding to the value of available sight lines in an environment (Braaksma *et al.*, 1980; Dada & Wirasinghe, 1999; Tosic & Babic, 1984), and the inter-connection density measures the complexity of a floor plan (O'Neill, 1991). The WBNM (Farr *et al.*, 2014) combined both the human and environmental aspects into one BN model and investigated the joint influence of these factors on effective wayfinding in airports.

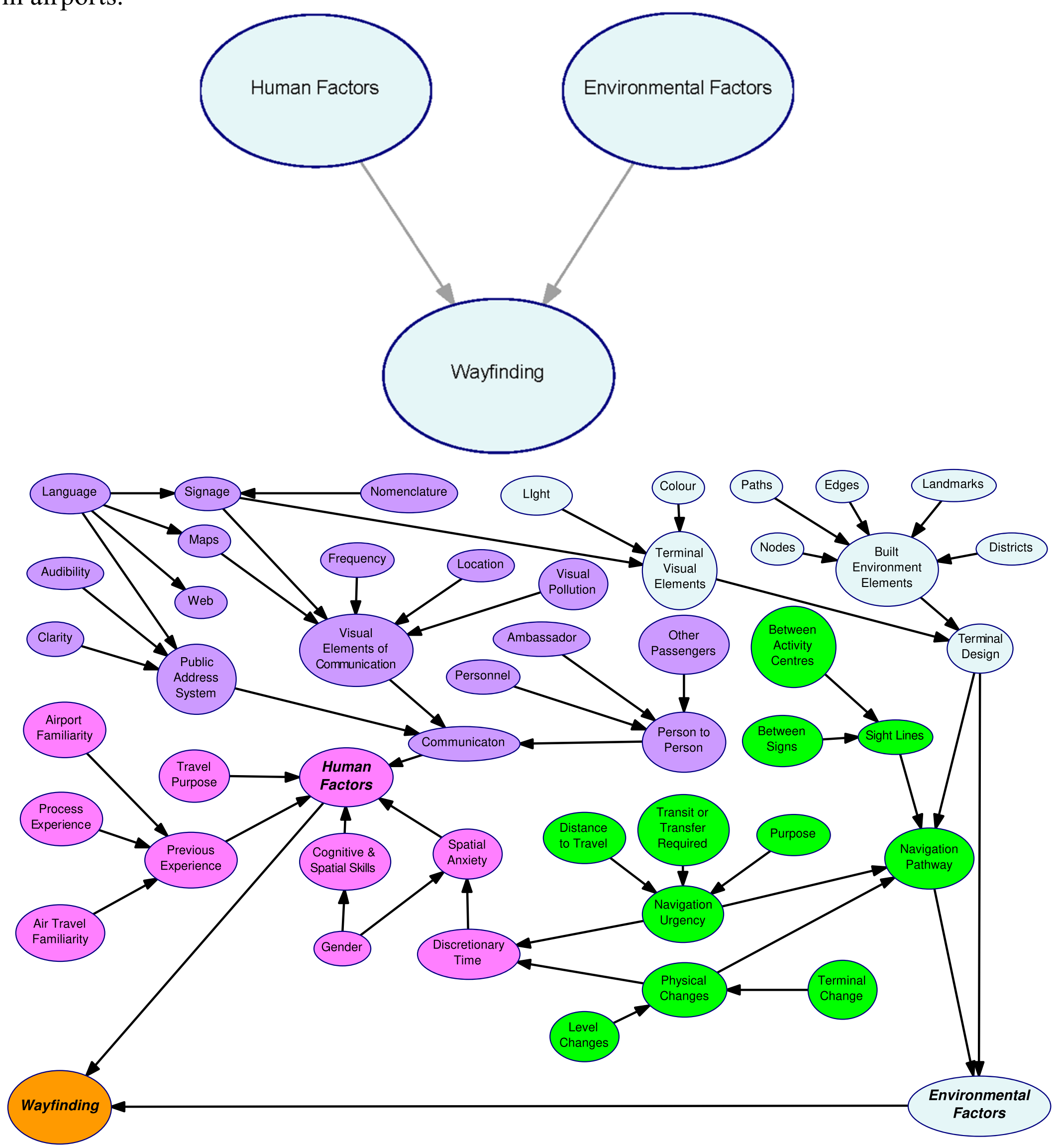


Figure 1. *The wayfinding Bayesian network model comprising three interconnected subnetworks (top) that expand to a full model (bottom)* (Farr *et al.*, 2014) . *[Colour figure can be viewed at wileyonlinelibrary.com]*

The model was developed via focus groups composed of a multidisciplinary team with differing levels of air travel and airport experience, by a review of the wayfinding research (Farr *et al.*, 2012) and from feedback from an audience of airport operators and BN modellers (Farr *et al.*, 2014). It was quantified using a combination of data obtained from a focus group using the Delphi method (Landeta, 2006), literature on wayfinding and an online survey. This survey was re-released to obtain more participants for this study, and the results from the 99 respondents were used (their values are provided in the Appendix). The WBNM, shown in Figure 1, is composed of five interconnected subnetworks, but we will concentrate just on three of them: one (pink) that includes the human factors, one (light blue) that includes the environmental factors and one (orange) that connects these two subnetworks to the target node representing the probability of effective wayfinding. The full network comprises 49 nodes and 58 connections. The analysis described in this paper focuses on the subnetwork comprising the three primary nodes, namely, human factors, environmental factors and wayfinding.

3 Bayesian Networks

A BN is a graphical representation of the joint probability distributions of a set of variables (Pearl, 1985; 1986) and is used for reasoning under uncertainty. A BN represents variables of interest as nodes and the dependencies between the variables as arcs. The variables underlying the nodes of the BN can be continuous, ordered or categorical; or alternatively continuous variables can be discretised to allow for ease of elicitation and computation (Korb & Nicholson, 2010). Examples of common discrete nodes are Boolean, ordered values, integer values and ranges of values. The number of categories is usually chosen in light of the context, desired inferences, available information and computational complexity. For example, all of the nodes in the WBNM are binary, as described in Farr *et al.* (2014).

A BN with binary nodes is depicted in Figure 2. It is immediately obvious that this representation is equivalent to a DAG, composed of nodes representing the variables of interest, arcs that show the direct influences between these variables, prior probability tables for the nodes that have no parents and conditional probability tables for the other nodes (Valtorta & Huang, 2008), like in Figure 2(c).

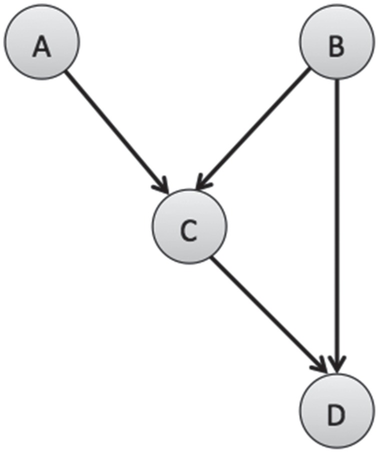
More precisely, for a DAG given by $\mathcal{G} = (\mathrm{V}, \mathrm{E})$, where V is the set of nodes and E is the set of directed links between nodes, a joint probability distribution $\mathrm{P}(\mathbf{X}_\mathrm{V})$ over the set of variables $\mathbf{X}_\mathrm{V}$ can be factorised as

$$\mathrm{P}(\mathbf{X}_\mathrm{V}) = \prod_{v \in \mathrm{V}} \mathrm{P}\left(\mathbf{X}_\mathrm{V} | \mathbf{X}_{\mathrm{pa}(v)}\right), \quad (1)$$

where $\mathbf{X}_{pa(v)}$ is the set of parent variables of variable $\mathbf{X}_\mathrm{V}$ for each node $v \in \mathrm{V}$. This provides the defining property of a BN; that is, the joint distribution of a node is conditioned only on the parents of that node.

For the network shown in Figure 2(a) with the states listed in Figure 2(b), the joint probability table given by $\mathrm{P}(A, B, C, D)$ would have $2^4 = 16$ entries. However, the constraints implied by the conditional independence structure in Figure 2(a) lead to $\mathrm{P}(A, B, C, D) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C|A, B)\mathrm{P}(D|B, C)$, which only contains $1 + 1 + 4 + 4 = 10$ parameters.

Some terminologies used in the BN literature and illustrated in Figure 2(a) are as follows (Korb & Nicholson, 2010). First, a node is a parent of a child if an arc goes from the former to the latter; for example, nodes A and B are the parents of C, and nodes C and D are the children of B. Second, if a directed chain of nodes exists, one node is an ancestor of another if it appears earlier in the chain, and it is a descendant of another if it comes later in the chain; for example,



(a)

Node	States
A	true, false
B	true, false
C	true, false
D	true, false

(b)

	A = true		A = false	
	B = true	B = false	B = true	B = false
C = true	0.45	0.2	0.6	0.77
C = false	0.55	0.8	0.4	0.23

(c)

Figure 2. *(a) A sample Bayesian network, with four nodes of interest, (b) the states of each of the variables and (c) the underlying conditional probability table for node* $\boldsymbol{C}$*, given nodes* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*.*

node D is a descendant of A. Third, a node without parents is a root node, for example, node A. Finally, a node without children is a leaf node, for example, node D.

The BN representation also allows for a reduction in the time needed to compute the marginal probabilities, which is the most common operation undertaken on a BN (Pearl, 1986; Valtorta & Huang, 2008). When new knowledge is obtained, beliefs are updated in a straightforward manner (Lauritzen & Richardson, 2002). Software such as Hugin (`www.hugin.com`), GeNIe & SMILE (`www.bayesfusion.com`) and Netica (`www.norsys.com/netica.html`) are able to perform these updates in an efficient manner.

4 Current Approaches for Combining Information in Bayesian Networks

Using opinions from multiple sources or experts to parameterise a BN is standard practice, particularly in situations where data are not available. This, however, raises the problem of *how* these opinions should be combined and used in a BN. Linear pooling (McConway, 1981) is a common way of combining the probabilities obtained from multiple experts or sources. The probability of an event X is approximated by averaging n conditional probabilities $P(X|\mathbf{E}_i) = P_i(X)$, provided by different sources of information or experts $\mathbf{E}_i$, $1 = 1, \ldots, n$, without knowing the joint model given by $P(P(X|\mathbf{E}_1), \ldots, P(X|\mathbf{E}_n))$. The probabilities in question are calculated by

$$P(X) = \sum_{i=1}^{n} \lambda_i P_i(X), \tag{2}$$

where λ_i are positive weights given to each of the n experts and $\sum_{i=1}^{n} \lambda_i = 1$.

Although the weights λ_i can sometimes be determined empirically given suitable data, they are often prescribed *a priori* on the basis of the problem-specific context. In the case study considered in this paper, each expert is given equal weighting, based on the premise that the wayfinding process is quite a person-specific experience, so all experiences were considered equally valuable. Thus, $\lambda_i = 1/n$ and

$$P(X) = \sum_{i=1}^{n} P_i(X)/n.$$

Two kinds of linear pooling can be used to combine expert opinions for use in BN models (Farr *et al.*, 2018). Prior linear pooling describes the process by which elicited probabilities are pooled within each node and the resultant conditional probability tables are then propagated through the network to find the marginal probabilities for the nodes of interest. In contrast, posterior linear pooling describes the process of quantifying and computing the BN for each expert separately, and the marginal probability distributions for the final nodes in the n BNs are then pooled. Thus, in prior linear pooling, Equation (2) is applied to each node separately in order to combine the opinions provided by the n experts into a pooled probability table for that node, whereas in posterior linear pooling, Equation (2) is applied to combine the n marginal probability tables for the final nodes of the BN.

Despite the conceptual simplicity of linear pooling, there are some serious drawbacks to this approach (Genest & Zidek, 1986). First, pooling only gives a point estimate for the consensus, losing the variety of opinions across the experts. Second, pooling, particularly when used with BNs, does not follow from a coherent probability model (de Finetti, 1964). That is, linear pooling can be considered an estimator if each observation is normally distributed and independent. The posterior mean is the sample average, and this implies that the errors are normally distributed. However, in the cases considered here, the data are discrete and often binary, so they cannot be generated by a normal distribution (without substantive assumptions), and no coherence can exist. Hence, it follows that linear pooling cannot follow from a coherent probability model. Third, as illustrated in the case study below, the different linear pooling methods can result in different outcomes for the nodes of interest. Finally, the conditional independence structure of the BN is not reflected in the way in which the expert opinions are combined, particularly in the case of prior linear pooling.

As an example of the last point, node X influences node Y, and each node has the states 'T' and 'F'. If n experts provide their opinions, there would be n probabilities for $P(X = \mathrm{T})$, $P(X = \mathrm{F})$, $P(Y = \mathrm{T}|X = \mathrm{T})$, $P(Y = \mathrm{T}|X = \mathrm{F})$, $P(Y = \mathrm{F}|X = \mathrm{T})$ and $P(Y = \mathrm{F}|X = \mathrm{F})$. By applying prior linear pooling, the average for each of these probabilities is found; however, these are not a reflection of what was originally given by the experts. That is, when the initial expert opinions were obtained, the four probabilities, $P(Y = \mathrm{T}|X = \mathrm{T})$, $P(Y = \mathrm{T}|X = \mathrm{F})$, $P(Y = \mathrm{F}|X = \mathrm{T})$ and $P(Y = \mathrm{F}|X = \mathrm{F})$, were given as conditional probabilities. By pooling these probabilities, the conditional independence structure is lost. This is illustrated in Table 1 where a toy example is shown. Here, the pooled probabilities from five experts are combined via prior linear pooling in a BN. These values can be markedly different from the original probabilities given by the experts questioned, particularly if experts have very different opinions. Considering $P(Y = \mathrm{T})$, for example, prior pooling gives a value of 0.5944, whereas 0.62 is obtained when using posterior pooling.

Table 1. *Example: comparison among elicited and pooled probabilities in a BN, based on five experts* $\mathbf{E}_i, i = 1, \ldots, 5$.

	$\mathbf{E}_1$	$\mathbf{E}_2$	$\mathbf{E}_3$	$\mathbf{E}_4$	$\mathbf{E}_5$	**Pooled value**
$P(X = \mathrm{T})$	0.8	0.7	0.6	0.8	0.5	0.68
$P(X = \mathrm{F})$	0.2	0.3	0.4	0.2	0.5	0.32
$P(Y = \mathrm{T} \mid X = \mathrm{T})$	0.6	0.7	0.6	0.8	0.4	0.62
$P(Y = \mathrm{F} \mid X = \mathrm{T})$	0.4	0.3	0.4	0.2	0.6	0.38
$P(Y = \mathrm{T} \mid X = \mathrm{F})$	0.5	0.3	0.6	0.5	0.8	0.54
$P(Y = \mathrm{F} \mid X = \mathrm{F})$	0.5	0.7	0.4	0.5	0.2	0.46
$P(Y = \mathrm{T})$	0.58	0.58	0.6	0.74	0.6	0.594

An alternative approach, proposed in this paper, is to consider a measurement error model for combining expert opinions for use in BNs. The approach uses the posterior probabilities ascribed to each node in the BN, which are computed from the prior information given by each expert. These observed probabilities are assumed to be noisy 'measurements' of the true probabilities and allow the representation of the systematic variation due to experts. This is described in more detail in Section 7.

5 Measurement Error Models

In almost all fields of statistical modelling, there is measurement error in the data generating process, such that the observations of a variable X vary from the underlying true value. The reasons for this variation can include inaccuracies in the recording device, potential bias or misclassification due to the study design, data collection practicalities in observational or experimental studies, and errors in data input. These errors can induce random or systematic variation, and, if ignored, the parameter estimates and confidence intervals in statistical models can suffer from serious biases (Muff *et al.*, 2014). If information about the errors induced in the measurement process is available, it may be useful to include them directly into the model. Information derived from experts can also be seen as a form of measurement error, in that the quantitative information elicited from each expert can be considered as a noisy realisation of an underlying common value.

There is a large literature on frequentist methods for addressing measurement error in regression (Carroll *et al.*, 1999; Carroll *et al.*, 2006; Gustafson, 2003) and a growing literature on Bayesian methods for this issue (Muff *et al.*, 2014; Richardson & Gilks, 1993; Stephens & Dellaportas, 1992). Bayesian approaches provide a natural framework for the inclusion of measurement errors, because discrepancies between observed and true values of a variable can be considered as, and described through, prior distributions, and the linear predictors can then be written in terms of the true values. In addition, a Bayesian model has been argued to be more straightforward to implement via Markov chain Monte Carlo (MCMC) than analogous frequentist models via expectation–maximisation (Carroll *et al.*, 2006).

6 Generalised Linear Models: A Baseline Case

A GLM extends linear regression by relating the linear model to the response variable via a link function and allowing the magnitude of the variance of each measurement to be a function of its predicted value (Nelder & Wedderburn, 1972). Assuming that there are n observations in a GLM, the data would be $(\mathbf{y}, \mathbf{z}, \mathbf{x})$, where the response variable is given by $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, the covariate matrix of dimension $n \times p$ for p error-free covariates is given by $\mathbf{z} = (z_1, \ldots, z_p)$

and $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ is the single error-prone covariate whose true values are unobservable. In the case study, this can be applied to the n expert opinions used to quantify the BN. Generalisation to multiple error-prone covariates can be achieved by assuming that $\mathbf{y}$ comes from the exponential family with mean given by $\mu_i = E(y_i|x_i)$ and is linked to the linear predictor η_i via

$$\begin{aligned} \mu_i &= h(\eta_i), \\ \eta_i &= \beta_0 + \beta_x x_i + \mathbf{z}_{[i,]z}, \end{aligned} \tag{3}$$

where $h(\cdot)$ is a known response function, β_0 is the intercept, β_x is the fixed effect for the error-prone covariate $\mathbf{x}$ and $\mathbf{z}_{[i,]}$ is a $1 \times p$ vector with corresponding vector of fixed effects given by β_z. By letting $\mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{T}}$ be the observed version of the true, but unobservable, covariate $\mathbf{x}$, it is possible to formulate the classical measurement error model.

6.1 Classical Measurement Error Model

The classical measurement error model assumes that the covariate $\mathbf{x}$ can only be observed by a proxy $\mathbf{w}$ such that $\mathbf{w} = \mathbf{x} + \mathbf{u}$. The error vector is given by $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$, the components of which are assumed to be independent and normally distributed with mean $= 0$ and common variance τ_u^{-1} (Muff *et al.*, 2014). In a regression set-up, if the error term $\mathbf{u}$ is assumed to be independent of the true covariate $\mathbf{x}$, any of the other covariates $\mathbf{z}$ and the response $\mathbf{y}$, then $\mathbf{y}$ and $\mathbf{w}$ are conditionally independent given $\mathbf{z}$ and $\mathbf{x}$. This means that given the true covariate $\mathbf{x}$ and covariates $\mathbf{z}$, having $\mathbf{w}$ provides no further information about the response variable, $\mathbf{y}$ (Carroll *et al.*, 2006; Muff *et al.*, 2014).

6.2 Measurement Error Models and Bayesian Networks

To our knowledge, there is no published literature on using measurement error models to combine expert opinions in BNs. The closest work has been by Marella & Vicard (2013), in which a mixed measurement error model was proposed and an object-oriented BN (OOBN) (Koller & Pfeffer, 1997) framework was used to implement the model. OOBNs are an extension of BNs where, instead of a node representing only a variable of interest, it can also contain nodes that are instances of other networks. The OOBN paradigm allows for hierarchical definition and construction of a BN by using network classes. The measurement model proposed in Marella & Vicard (2013) describes the relationship between observed and true categories in a questionnaire survey. The results from the measurement model are then used in an OOBN, which is then used to represent, in a single model, the entire survey process.

7 Proposed Model

For exposition and without loss of generality, consider a BN in which each node is binary, so that the information provided by the expert is in the form of the probability associated with one of the outcomes of the node. A measurement error model is proposed to treat the experts' probabilities for a node as observations of the underlying true probabilities, subject to systematic variation. While the focus of this work is primarily on the situation where consensus or agreement is formed simultaneously for multiple nodes in a BN, it is noted that a random effects model can be applied when considering only a single node.

In order to pool the probabilities on the correct 0 to 1 scale, a Beta distribution for the response variable is used, because it is more suitable as a data generating mechanism than the Gaussian error distribution implied by the linear pooling estimator. This is similar to the work of Ferrari & Cribari-Neto (2004) and Figueroa-Zúñiga *et al.* (2013), where the response variable is assumed to be beta distributed with the mean and the precision parameter modelled using fixed and random effects.

For the univariate model, where the consensus is formed for a single node of interest, consider

$$p_i \sim \text{Beta}(a_i, b_i), \tag{4}$$

where p_i is the marginal probability for expert $i = 1, \ldots, n$. To allow for variation between experts, we take

$$\begin{aligned} \text{logit}\left(\frac{a_i}{a_i + b_i}\right) = \text{logit}\left(\frac{a_i}{b_i}\right) &= \mu + \epsilon_i, \\ \text{where} \quad \mu &\sim \text{N}(0, \tau_\mu^{-1}), \\ \epsilon_i &\sim \text{N}(0, \tau_\epsilon^{-1}), \end{aligned}$$

where the hyperparameters τ_μ and τ_ϵ are specified according to the problem.

As the expected value of the logit term equals 0, this implies that $a_i = b_i$, so an alternative construction is to consider a distribution symmetric around $p_i = 1/2$ and impose a prior on $a_i + b_i$,

$$a_i + b_i \sim \text{Gamma}(\alpha_0, \beta_0), \tag{5}$$

where, as above, the hyperparameters α_0 and β_0 are problem specific.

An analogous multivariate measurement error (MME) model can be developed when forming consensus for multiple nodes in a BN. Consider

$$p_{ij} \sim \text{Beta}(a_{ij}, b_{ij}), \tag{6}$$

where p_{ij} is the marginal probability for expert $i = 1, \ldots, n$ at node $j = 1, \ldots, m$. A multivariate Gaussian random effect for each expert can be used because the probabilities ascribed by experts to each node are treated as observations of the underlying true probabilities, which are closer around a mean value. To allow for extra variation due to the heterogeneity between experts, an independent random effect $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_{i1}, \ldots, \boldsymbol{\epsilon}_{im})$ for each expert i was added and a vague normal prior given to the mean, so

$$\begin{aligned} \text{logit}\left(\frac{a_{ij}}{a_{ij} + b_{ij}}\right) &= \mu_j + \boldsymbol{\epsilon}_{ij}, \\ \text{where} \quad \mu_j &\sim \text{N}(0, \tau_\mu^{-1}), \\ \boldsymbol{\epsilon}_i &\sim \text{N}\left(\mathbf{0}_m, \mathbf{Q}^{-1}\right), i = 1, \ldots, n, \\ a_{ij} + b_{ij} &\sim \text{Gamma}(\alpha_\tau, \beta_\tau). \end{aligned} \tag{7}$$

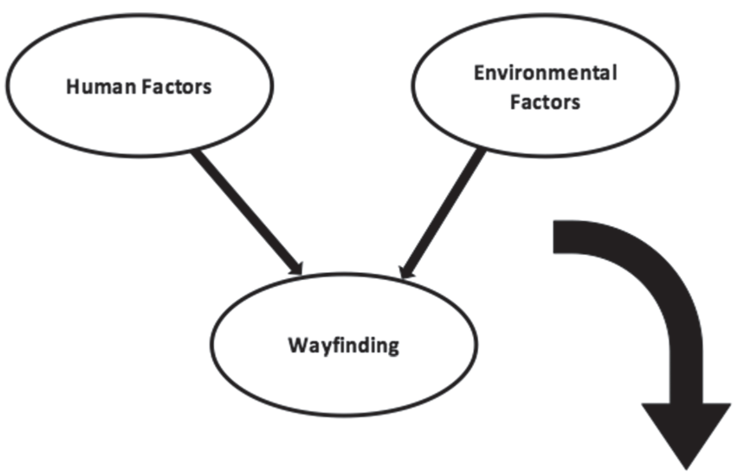
The structure of the random effect term, $\boldsymbol{\epsilon}$, is such that $\boldsymbol{\epsilon} = \mathbf{Rs}$, where $\mathbf{R}$ is the Cholesky decomposition of the precision matrix $\mathbf{Q}$ and $\mathbf{s}$ is a vector of i.i.d. standard normals, that is, $\mathbf{s} = \text{N}(0, I)$. By definition, if $\boldsymbol{\epsilon} = \mathbf{Rs}$ for $\mathbf{s}$ i.i.d. normals, then $\boldsymbol{\epsilon}$ has the precision matrix such that $\mathbf{Q} = (\mathbf{RR}^T)^{-1}$ (Eaton, 2007). This implies that if $\mathbf{R}$ has the correct sparsity required, then $\mathbf{Q}$ will also have the correct sparsity structure (Rue & Held, 2005). This is an indirect way in which to put a prior on precision matrices with a fixed sparsity structure, or equivalently, on Gaussian distributions with the right conditional independence structure. By applying this to the multivariate model, it follows that $\boldsymbol{\epsilon} \sim \text{N}(\mathbf{0}_m, \mathbf{RR}^T)$. Hence, similarly, by finding $\mathbf{R}$, we are able to give $\mathbf{Q}$ the right structure that reflects the conditional independence of a BN.

8 The Three-node Multivariate Measurement Error Model

In this section, we illustrate the use of a measurement error model for combining experts' opinions in a three-node BN corresponding to the subnetwork of the WBNM described in Section 2. This provides sufficient opportunity to demonstrate the feasibility and utility of the approach in a simple, but significant, case. Recall that the subnetwork is structured as two root parent nodes, environmental factors (E) and human factors (H), connected by directed arcs to the child leaf node of wayfinding (W), as depicted in Figure 3.

Following the model described in the previous section, we observe that the precision matrix $\mathbf{Q}$ of the random effect, $\boldsymbol{\epsilon}_i$, gives information about the conditional probabilities and ensures that the expert opinions flow through the BN (here from human and environmental factors to wayfinding), as shown in (8). The proposed model allows also for the combination of all the opinions, so that all of them will count. If reordering of the independent expert opinions occurs, coherence is maintained and there will not be an impact on the result of the model. Additionally, $\mathbf{Q}$ allows the model to 'borrow strength' from other parts of the model (Tukey, 1974). That is, information from one node can be used to inform other nodes as information is able to travel up and down the levels of the hierarchy (Efron, 2010). The issue of ensuring that the conditional independence structure of the BN is reflected when combining expert opinions is addressed by the precision matrix $\mathbf{Q}$. In order for the consensus model to be consistent with the BN structure, the conditional independence structure of the Gaussian random effect was forced to mirror that of the BN. This forces a sparsity structure on the precision matrix $\mathbf{Q}$, such that $\mathbf{Q}_{ij} \neq 0$ iff node i depends on node j in the BN.

The structure of the precision matrix, $\mathbf{Q}$, is constructed to reflect the conditional independence structure of these nodes, as shown in Figure 3.



$$Q = \begin{pmatrix} \times & & \times \\ & \times & \times \\ \times & \times & \times \end{pmatrix}$$

Figure 3. *The sparsity structure of the precision matrix,* $\mathbf{Q}$. *The structure of this matrix reflects the conditional independence structure of the final three nodes of the wayfinding Bayesian Network model.*

This requires the construction of **R**, the Cholesky decomposition of **Q**. This is where the expert priors on **Q** are used. Recall that in (6), $\text{logit}\left(\frac{a_{ij}}{a_{ij}+b_{ij}}\right) = \mu_j + \boldsymbol{\epsilon}_{ij}$. It is possible to write $\boldsymbol{\epsilon}_i = \text{Rs}$, which leads to

$$\left.\begin{aligned} \text{logit}\mu_{\text{H}} &= \mu_{\text{H}} + \beta_1\epsilon_{\text{W}} + \epsilon_{\text{H}} \\ \text{logit}\mu_{\text{E}} &= \mu_{\text{E}} + \beta_2\epsilon_{\text{W}} + \epsilon_{\text{E}} \\ \text{logit}\mu_{\text{W}} &= \mu_{Q} + \beta_3\epsilon_{\text{H}} + \beta_4\epsilon_{\text{E}} + \epsilon_{\text{W}} \end{aligned}\right\}, \tag{8}$$

where $\mu_{\text{H}}, \mu_{\text{E}}$ and μ_{W} are the mean opinions for nodes H, E and W, respectively; and $\epsilon_{\text{H}}, \epsilon_{\text{E}}$ and ϵ_{W} are the random effects associated with the expert opinions for nodes H, E and W, respectively. The β terms give a measure of how much of the random effects comes from the other nodes. That is, β_1 says how much noise from W influences H, β_2 is a measure of how much W influences E and β_3 and β_4 give the size of the influence of H and E, respectively, on W.

In vector form, this gives

$$\text{logit}\boldsymbol{\mu}_X = \boldsymbol{\mu}_X + \text{Rs},$$

where **R** is given by

$$\mathbf{R} = \begin{pmatrix} \tau_{\text{H}}^{-1/2} & 0 & \beta_1\tau_{\text{W}}^{-1/2} \\ 0 & \tau_{\text{E}}^{-1/2} & \beta_2\tau_{\text{W}}^{-1/2} \\ 0 & 0 & \tau_{\text{W}}^{-1/2} \end{pmatrix},$$

with $\tau_X \sim \text{Gamma}(1, 5 \times 10^{-5})$ and $\beta_X \sim \text{N}(0, 5 \times 10^{-5})$. As $\mathbf{Q} = (\mathbf{RR}^T)^{-1}$ (Eaton, 2007), this gives the precision matrix as

$$\mathbf{Q} = \begin{pmatrix} \tau_{\text{H}} & 0 & -\beta_1\tau_{\text{H}} \\ 0 & \tau_{\text{E}} & -\beta_2\tau_{\text{E}} \\ -\beta_1\tau_{\text{H}} & -\beta_2\tau_{\text{E}} & \beta_1^2\tau_{\text{H}}\tau_{\text{W}}^2 + \beta_2^2\tau_{\text{E}}\tau_{\text{W}}^2 + \tau_{\text{W}} \end{pmatrix}.$$

It should be noted that if all of the βs are zeros, then the random effects $\epsilon_{\text{H}}, \epsilon_{\text{E}}$ and ϵ_{H} are independent and **Q** is a diagonal precision matrix.

9 Results

The MME model given by Equation (6) was applied to the final three nodes of the WBNM. As discussed in Section 2, the dataset used for the analysis (available in the Appendix) comprised the set of opinions elicited from the $n = 99$ experts about the states 'good', 'good' and 'effective' for the human factors, environmental factors and wayfinding nodes, respectively. A Gamma(1, 0.1) distribution was specified for the prior on the terms $a + b$, in line with the ambition to have proper but relatively uninformative priors. This was justified by calculating an approximate 95% interval for the anticipated values for p_i implied by this prior, obtained by taking the 2.5% and 97.5% quantiles for the prior, letting $a_i = b_i$ and undoing the logit transformation. The obtained interval was 0.53 to 1, which was considered to be reasonable. Proper but relatively uninformative priors were also specified for μ and ϵ, with $\tau_{\mu}^{-1} = 10^4$ producing a diffuse distribution on μ and hence an almost uniform distribution on p_i, and $(a_\tau, b_\tau) = (1, 5 \times 10^5)$ producing a distribution for ϵ that reflected a relatively small contribution of the measurement error to the overall value of p_i.

The analysis was undertaken using the R package integrated nested Laplace approximation (R-INLA) (Rue *et al.*, 2009). R-INLA allows full Bayesian inference to be performed on a class

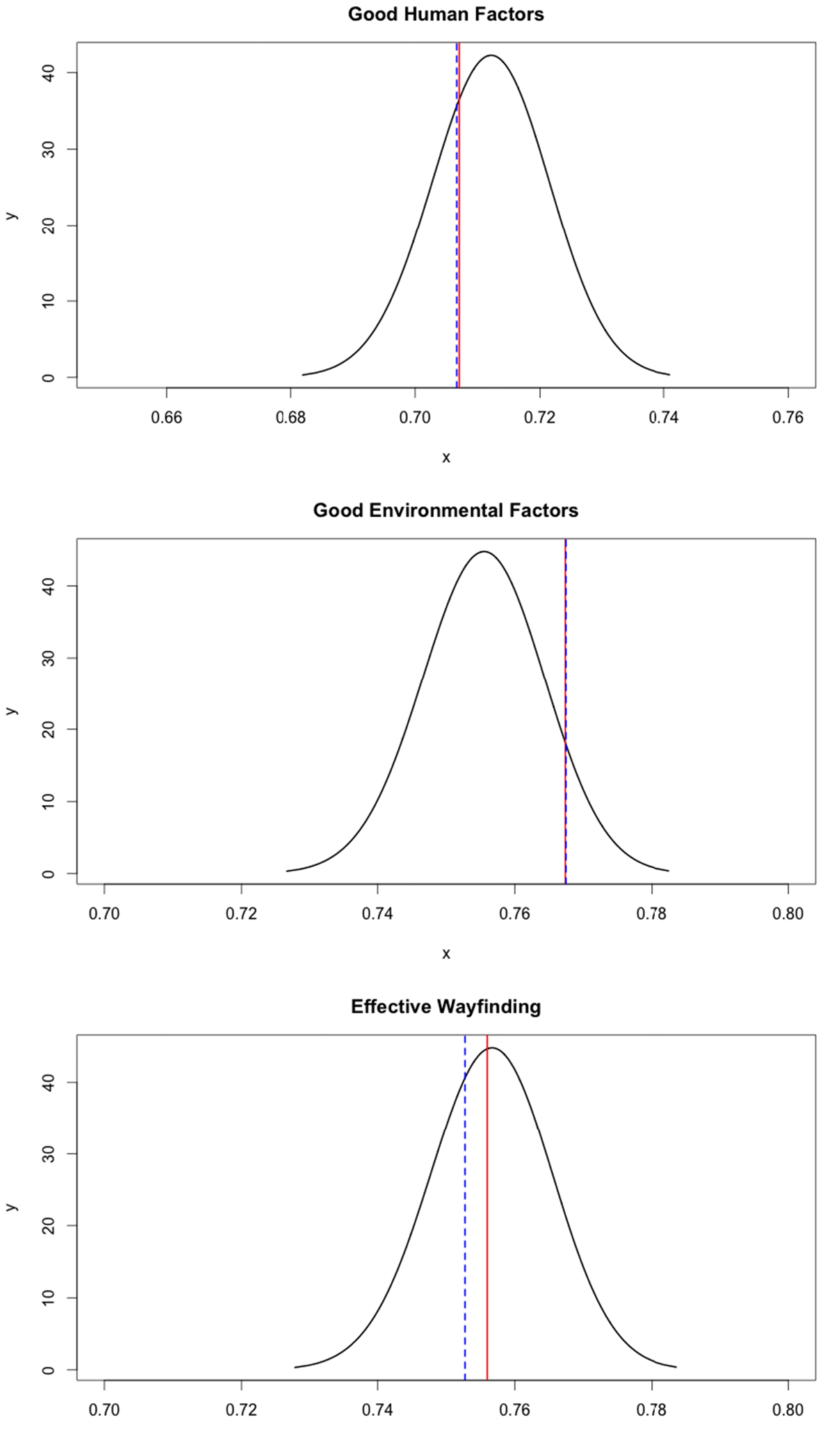


Figure 4. *The results of the implementation of the multivariate measurement error model (solid black line), prior linear pooling (dashed blue line) and posterior linear pooling (solid red line) on the final three nodes of the wayfinding Bayesian network model. [Colour figure can be viewed at wileyonlinelibrary.com]*

of latent Gaussian models including spatial models, geostatistical models, generalised linear mixed models and generalised additive models (Martins *et al.*, 2013). It utilises deterministic Laplace approximations by fitting Gaussian conditional posteriors via an optimisation step for latent Gaussian models. The approximation can also be used in a nested framework and provides a faster and more accurate alternative to simulation-based MCMC schemes (Martins *et al.*, 2013). The multivariate model given by (6) and its application to the final three nodes of the WBNM are a good candidate for using R-INLA owing to the sparsity of **Q** and the ability to perform fast, easy and accurate computation for this class of problem. The R-INLA code used for the analysis is provided in the Appendix.

The resulting distributions from the MME model are shown by the solid black line in Figure 4. For comparison, prior and posterior linear pooling were also performed using the same BN and dataset. The results are also shown in Figure 4, with the distributions depicted in blue and red, respectively.

The results for both pooling methods are almost identical for the human and environmental factor nodes at 0.708 and 0.768, respectively, with the MME model mean differing in both cases at 0.715 and 0.7587 for the respective nodes. The posterior pooling result and the MME model mean for the wayfinding node are almost the same at 0.755 and are different to the prior pooling result, which is 0.752. This may be because both the MME model and the posterior pooling follow the conditional independence structure of the BN, whereas the prior pooling method does not. By using the MME model, it is possible to obtain a distribution for the nodes of interest. This is more informative than a point estimate like that obtained using the pooling methods as it describes the uncertainty associated with the estimates.

To investigate if a small sample size has an impact on the results, the MME model and the posterior pooling were implemented for $n = 15$ randomly selected data points for the final three nodes of the WBNM. The distribution for the MME model is shown in Figure 5 as the solid black line, with the pooled results shown as the solid red line. For all three nodes, the MME model and the posterior pooling results were comparable. Given the choice of which method to use, the MME model would still be preferable to posterior pooling for the reasons noted earlier. That is, it follows from a coherent probability model that the resulting distribution is more informative than a point estimate, and it allows for uncertainty.

With the use of the same subset of 15 data points, the univariate model given by (4) was also implemented for each of the three final nodes of the WBNM. This method is easier to implement than the MME model because the conditional independence structure of **Q** does not have to be calculated. That is, the βs in Equation (8) are zero, and **Q** is a diagonal matrix. The results of the implementation of this model are shown by the dashed lines in Figure 5. For each of the three nodes, the probability mass for the MME model resulted in both better location and a more conservative spread than that of the univariate model. This is because the MME model is able to borrow strength from other nodes in order to better inform the model.

10 Discussion

The MME model proposed in this paper addresses the issues associated with the current pooling methods used for combining expert opinions in BNs. Namely, the issues are that pooling, when used with BNs, does not follow from a coherent probability model, the conditional independence structure of the BN is not followed, particularly in the case of prior pooling, and pooling only gives a point estimate for the consensus. The MME model overcomes these issues by using a measurement error model to treat each of the probabilities ascribed by experts to

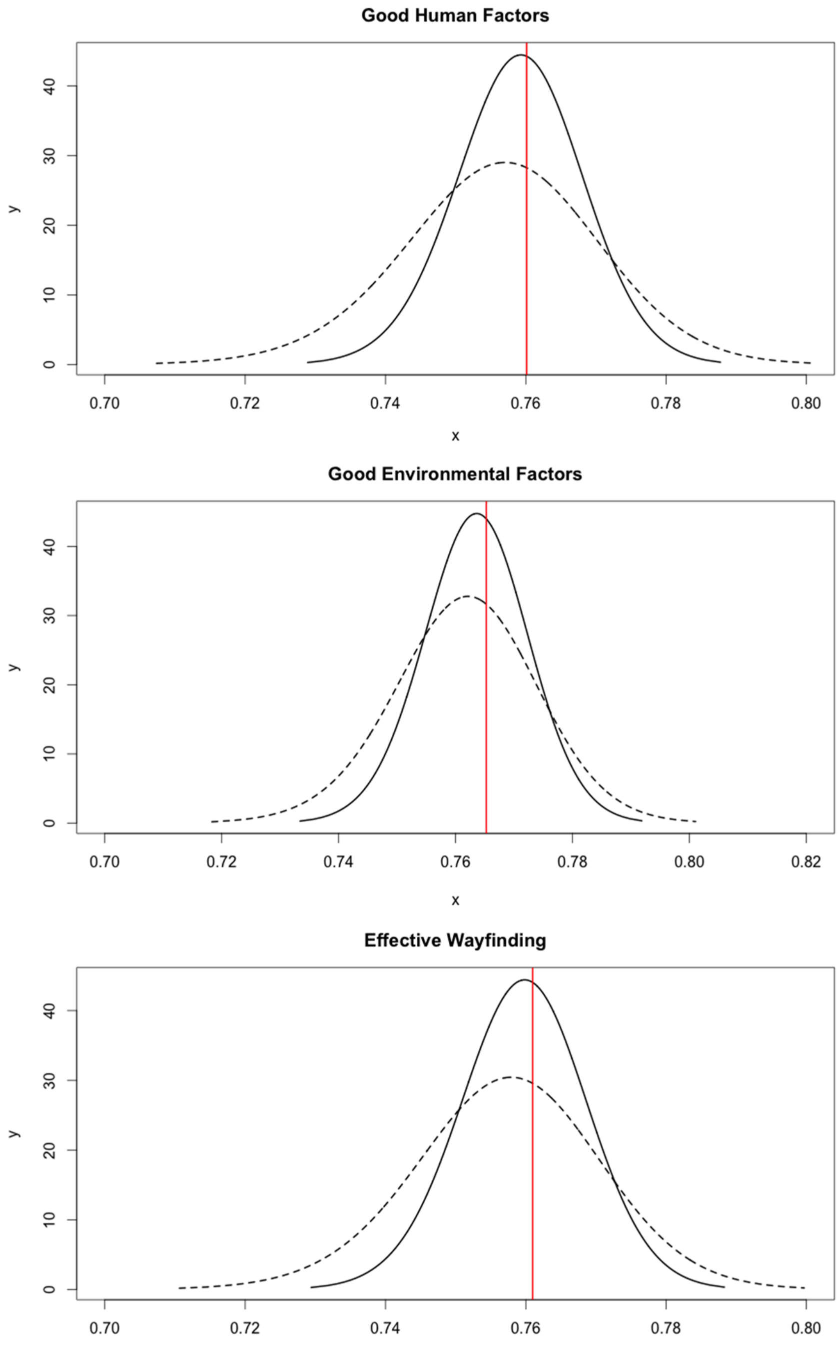


Figure 5. *The results from an implementation of the multivariate measurement error model (solid black line), posterior pooling (red line) and the univariate model applied to each node (dashed black line) on* $n = 15$ *randomly selected data points. [Colour figure can be viewed at wileyonlinelibrary.com]*

each node as observations of the underlying true probabilities that are subject to systematic variation due to experts.

The systematic variation is modelled through the random effect term, $\boldsymbol{\epsilon}_i$, which contains the precision matrix **Q**. By ensuring the correct structure and sparsity of **Q**, the conditional independence of the BN is reflected in the model. Additionally, **Q** also contains information regarding the conditional probabilities in the BN and as such addresses the issue of coherence. Finally, by using a measurement error approach, it is possible to obtain a distribution for each state and node of interest. This makes it possible to obtain a measurement with uncertainty for any marginal probability of interest rather than just a point estimate.

It must be noted that because expert opinions may not necessarily follow the measurement error model (e.g. the experts questioned may all be biased), this uncertainty cannot be taken on face value. Keith (1996), for example, argues against the aggregation of opinions or, at least, the over-aggregation of opinions. This is particularly relevant to situations where opinions may be biased or extreme, because the combined distributions of the opinions may bear no similarity to the true distribution of the probabilities.

The MME model can be modified to cater for issues such as expert weighting and outlier detection. Expert information can be differentially weighted in the BN through the weights λ_i in Equation (2) or by imposing a differential inflation factor on the variances of the expert-specific priors in Equations (4) and (6). In this case study, all experts were weighted equally. Outlier detection can be incorporated by considering the conditional predictive ordinate (CPO) values (Geisser, 1980), which is a tool for detecting observations that are fitted poorly by a given model. In this case, it would measure how well an expert is predicted from the other experts. The CPO expresses the posterior probability of observing the value i when the model is fitted to all data except i. Very low CPO values imply that i is an outlier and an influential observation (Gelfand, 1996).

The MME model could also be extended to include bias and additional covariates. In the case study, if there was an interest in investigating the effect of experienced (E) and inexperienced (I) travellers, the overall mean for node j, μ_j, in the model given by (6) could be modified as follows:

$$\mu_{\mathrm{I}} \sim \mathrm{N}(\mu_j - \delta_{\mathrm{I}}, \sigma_{\mathrm{I}}^2),$$
$$\mu_{\mathrm{E}} \sim \mathrm{N}(\mu_j + \eta_{\mathrm{E}}, \sigma_{\mathrm{E}}^2),$$

where μ_{I} and μ_{E} are the means for inexperienced and experienced travellers, respectively; σ_{I}^2 and σ_{E}^2 are the variances for inexperienced and experienced travellers, respectively; and δ_{I} and η_{E} represent the impact of inexperienced and experienced travellers, respectively, on the overall mean, μ_j.

As discussed earlier, there is a strong connection between GLMs and measurement error ones, and it could be worth fitting a generalised multivariate regression model and comparing the results with the ones in the current paper. We believe that the sparsity structure of the BNs makes our approach more suitable, but we leave the practical comparison to future work.

Finally, it has to be noted that the method proposed here to obtain the correct conditional independence structure in the Gaussian random effect is hard to extend to more complicated networks. If more than three nodes were to be investigated, the use of G-Wishart priors for **Q**, which are restrictions of Wishart random variables to the subspace of matrices with the correct sparsity structure, has to be undertaken (Lenkoski, 2013). Recent results have derived direct samplers for G-Wishart random variables, which makes them practical in this application (Lenkoski, 2013; Wang & Li, 2012). An extension to more than three nodes means that an MCMC scheme rather than INLA would be used to perform the inference.

In this paper, we have focused on a three-node sub-graph of the full wayfinding BN. The methods described in this paper would, theoretically, scale up to a problem of that size. As with many multivariate models, we are defeated by an explosion in the number of parameters. For a BN with n nodes and n_c connections, fitting the MME model proposed in Section 3 requires the estimation of $n + n_c$ precision parameters and n intercepts. This is significantly smaller than the number of precision parameters needed for the MME model without the conditional independence assumptions; for the full wayfinding network, the number of precision parameters is reduced from 1125 to 107. Unfortunately, with only 99 data points, it is not feasible to fit a model of this complexity to the current data set.

ACKNOWLEDGEMENTS

This research forms part of the work undertaken by the Airports of the Future (LP0990135) project, which is funded by the Australian Research Council Linkage Project scheme. The authors also acknowledge the contributions made by the many aviation industry stakeholders also involved in this project. More details on Airports of the Future and its participants can be found at `https://research.qut.edu.au/aotf/`.

References

Beliakov, G., Pradera, A. & Calvo, T. (2007). *Aggregation Functions—A Guide for Practitioners* Berlin: Springer-Verlag.

Braaksma, J., Asce, A. & Cook, W. (1980). Human orientation in transportation terminals. *J. Transp. Eng.*, **106**, 189–203.

Burgman, M., McBride, M., Ashton, R., Speirs-Bridge, A., Flanders, L., Wintle, B., Fidler, F., Rumpff, L. & Twardy, C. (2011). Expert status and performance. *PLoS ONE*, **6**(7), e22998.

Carroll, R., Roeder, K. & Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics*, **55**, 44–54.

Carroll, R., Ruppert, D., Stefanski, L. & Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd Boca Raton, FL: Chapman and Hall/CRC.

Cooke, R. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science* Oxford: Oxford University Press.

Dada, E. & Wirasinghe, S. (1999). Development of a new orientation index for airport terminals. *Transport. Res. Rec.*, **1662**, 41–47.

Dalkey, N. & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Manage. Sci.*, **9**, 458–467.

de Finetti, B. (1964). Foresight: its logical laws, its subjective scores. Wiley, New York, pp. 57–118.

Eaton, M. (2007). Multivariate Statistics: A Vector Space Approach. Institute of Mathematical Sciences, Beachwood, OH, pp. 103–131.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction* Cambridge: Cambridge University Press.

Farr, A., Kleinschmidt, T., Johnson, S., Yarlagadda, P. & Mengersen, K. (2014). Investigating effective wayfinding in airports: a Bayesian network approach. *Transport*, **29**, 90–99.

Farr, A., Kleinschmidt, T., Yarlagadda, P. & Mengersen, K. (2012). Wayfinding: a simple concept, a complex process. *Transp. Rev.*, **32**, 715–743.

Farr, A., Ruggeri, F. & Mengersen, K. (2018). Prior and posterior linear pooling for combining expert opinions: uses and impact on Bayesian networks. *Entropy*, **20**, 209–222.

Ferrari, S. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *J. Appl. Stat.*, **31**, 799–815.

Figueroa-Zúñiga, J., Arellano-Valle, R. & Ferrari, S. (2013). Mixed beta-regression: a Bayesian perspective. *Comput. Statist. Data Anal.*, **61**, 137–147.

French, S. (1985). Group Consensus Probability Distributions: A Critical Survey Bayesian Statistics, 2nd ed. (J.M. Bernardo, M.H. DeGroot, D.V. Lindley & A.F.M. Smith, eds.) North-Holland, Amsterdam, pp. 183–201.

French, S. (2011). Aggregating expert judgment. *Revista de la Real Academia de Ciencias Exactas, Fásicas y Naturales (España)*, **105**, 181–206.

Friedman, N., Linial, M., Nachman, I. & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *J. Comput. Biol.*, **7**, 601–620.

Gärling, T., Böök, A. & Lindberg, E. (1984). Cognitive mapping of large-scale environments the interrelationship of action plans, acquisition, and orientation. *Environ. Behav.*, **16**, 3–34.

Geisser, S. (1980). Discussion on "Sampling and Bayes' inference in scientific modelling and robustness" by G.E.P Box. *J. Roy. Statist. Soc. Ser. A*, **143**, 416–417.

Gelfand, A. (1996). Model Determination Using Sampling Based Methods. In *Markov Chain Monte Carlo in Practice*, Eds. W. Gilks, S. Richardson & D. Spiegelhalter, pp. 145–161. London: Chapman and Hall.

Genest, C. & Zidek, J. (1986). Combining probability distributions: a critique and an annotated bibliography. *Statist. Sci.*, **1**, 114–135.

Gustafson, P. (2003). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments* FL: Chapman & Hall/CRC: Boca Raton.

Heckerman, D. (1990). Probabilistic similarity networks. *Networks*, **20**, 607–636.

Johnson, S. (2009). *Integrated Bayesian Network Frameworks for Modelling Complex Ecological Issues*, Ph.D. thesis, School of Mathematical Sciences, Faculty of Science and Technology, Queensland University of Technology.

Keith, D. W. (1996). When is it appropriate to combine expert judgments? *Clim. Change.*, **33**, 139–143.

Koller, D. & Pfeffer, A. (1997). Object-oriented Bayesian networks. In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence*. RI, Providence, pp. 302–313.

Korb, K. & Nicholson, A. (2010). *Bayesian Artificial Intelligence*, 2nd Boca Raton: FL: Chapman and Hall/CRC.

Kuipers, B. (1978). Modeling spatial knowledge. *Cogn. Sci.*, **2**, 129–153.

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technol. Forecast. Soc. Change*, **73**, 467–482.

Lauritzen, S. & Richardson, T. (2002). Chain graph models and their causal interpretation. *J. Royal Stat. Soc., Series B: Statistical Methodology*, **64**, 321–361.

Lenkoski, A. (2013). A direct sampler for g-Wishart variates. *Stat*, **2**, 119–128.

Lindley, D. (1983). Reconciliation of probability distributions. *Oper. Res.*, **31**, 866–880.

Marella, D. & Vicard, P. (2013). Object-oriented Bayesian networks for modeling the respondent measurement error. *Comm. Statist. Theory Methods*, **42**, 3463–3477.

Martins, T., Simpson, D., Lindgren, F. & Rue, H. (2013). Bayesian computing with INLA: new features. *Comput. Statist. Data Anal.*, **67**, 68–83.

McConway, K. (1981). Marginalization and linear opinion pools. *J. Amer. Statist. Assoc.*, **71**, 410–414.

Muff, S., Riebler, A., Held, L., Rue, H. & Saner, P. (2014). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *J. Roy. Statist. Soc. Ser. C*, **64**, 231–252.

Nelder, J. & Wedderburn, R. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A*, **135**, 370–384.

O'Neill, M. (1991). Effects of signage and floor plan configuration on wayfinding accuracy. *Environ. Behav.*, **23**, 553–574.

Passini, R. (1981). Wayfinding: a conceptual framework. *Urban Ecol.*, **5**, 17–31.

Passini, R. (1984). *Wayfinding in Architecture* New York: Van Nostrand Reinhold Company.

Pearl, J. (1985). Bayesian networks: a model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*. University of California, Irvine CA, pp. 329–334.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artif. Intell.*, **29**, 241–288.

Peponis, J., Zimring, C. & Choi, Y. (1990). Finding the building in wayfinding. *Environ. Behav.*, **25**, 555–590.

Pollino, C., Woodberry, O., Nicholson, A., Korb, K. & Hart, B. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environ. Model. Softw.*, **22**, 1140–1152.

Richardson, S. & Gilks, W. (1993). Conditional independence models for epidemiologcial studies with covariate measurement error. *Stat. Med.*, **12**, 1703–1722.

Rue, H. & Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications* Boca Raton: FL: Chapman and Hall/CRC.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J. Royal Stat Soc., Series B: Statistical Methodology*, **71**, 319–392.

Stephens, D. & Dellaportas, P. (1992). Bayesian analysis of generalised linear models with covariate measurement error. In *Bayesian Statistics*, 4th ed. (J.M. Bernardo, J. Berger, A. Dawid & A.F.M. Smith, eds.) Oxford University Press, Oxford, pp. 813–820.

Taroni, F., Biedermann, A., Garbolino, P. & Aitken, C. (2004). A general approach to Bayesian networks for the interpretation of evidence. *Forensic Sci. Int.*, **139**, 5–16.

Timpf, S., Volta, G., Pollock, D. & Egenhofer, M. (1992). A conceptual model of wayfinding using multiple levels of abstraction. In *Theories and Methods of Spatio-temporal Reasoning in Geographic Space* (A. Frank, I. Campari & U. Formentini, eds.) Springer-Verlag, Berlin, pp. 348–367.
Tosic, V. & Babic, O. (1984). Quantitative evaluation of passenger terminal orientation. *J. Adv. Transport.*, **18**, 279–295.
Trucco, P., Cagno, E., Ruggeri, F. & Grande, O. (2008). A Bayesian belief network approach for modelling human and organisational factors risk analysis: a case study in maritime industry. *Reliab. Eng. Syst. Safe.*, **93**, 845–856.
Tukey, J. (1974). Robustness study analysis (a further analysis of the first phase of the Princeton Robustness Study; examples of less standard two-way table analysis). In *Exploring Data Analysis: The Computer Revolution in Statistics* (W.J. Dixon & W.L. Nicholson, eds.) University of California Press, Berkeley, pp. 229-294.
Valtorta, M. & Huang, Y. (2008). Identifiability for causal Bayesian networks: a gentle introduction. *Cybern. Syst.*, **39**, 425–442.
Wang, H. & Li, S. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electron. J. Stat.*, **6**, 168–198.
West, M. (1988). Modelling expert opinion. In *Bayesian Statistics*, 3rd ed. (J.M. Bernardo, M.H. DeGroot, D.V. Lindley & A.F.M. Smith, eds.) Oxford University Press, Oxford, pp. 493–508.
Winkler, R. (1968). The consensus of subjective probability distributions. *Manage. Sci.*, **15**, B61–B75.

[*Received December 2017, accepted July 2019*]

Appendix: Dataset for the Three Nodes of the Wayfinding Bayesian Network Model

The following dataset was used for the case study analysis. The values represent the probabilities assigned by each of the 99 experts to the three nodes, environmental factors (ef), human factors (hf) and wayfinding (wf).

- **ef**: .77., .77., .77., .77., .77., .76., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .76., .77., .77., .76., .77., .77., .77., .77., .77., .77., .77., .77., .78., .77., .75., .76., .75., .78., .77., .76., .77., .77., .77., .77., .75., .77., .77., .77., .76., .76., .77., .76., .77., .77., .77., .77., .76., .77., .76., .77., .77., .77., .76., .76., .77., .76., .76., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .77., .76., .77., .77., .77., .77., .77., .77., .77., .77., .77., .76., .76., .76., .77., .76., .77., .76., .77., .77., .76

- **hf**: .71., .67., .75., .77., .42., .56., .74., .84., .75., .66., .72., .51., .69., .72., .47., .77., .71., .65., .73., .76., .66., .69., .78., .74., .68., .78., .79., .70., .74., .77., .84., .38., .72., .66., .84., .77., .77., .35., .77., .76., .81., .68., .77., .77., .78., .70., .25., .38., .78., .78., .57., .75., .74., .59., .78., .76., .78., .53., .39., .77., .80., .88., .83., .96., .71., .78., .77., .78., .58., .77., .44., .84., .78., .39., .64., .84., .77., .54., .80., .77., .81., .90., .86., .83., .63., .78., .89., .65., .83., .75., .74., .57., .67., .81., .35., .84., .82., .79., .74., .70

- **wf**: .65., .65., .58., .76., .83., .75., .70., .80., .71., .81., .71., .79., .84., .83., .69., .77., .76., .65., .71., .76., .79., .72., .78., .83., .72., .78., .82., .81., .71., .82., .78., .72., .83., .71., .83., .71., .70., .69., .60., .68., .88., .67., .79., .91., .59., .61., .91., .87., .74., .74., .88., .73., .81., .85., .63., .79., .63., .83., .95., .72., .75., .74., .70., .72., .84., .64., .83., .74., .79., .77., .80., .75., .83., .80., .86., .73., .75., .80., .62., .75., .58., .81., .82., .83., .92., .75., .73., .68., .74., .61., .70., .91., .95., .71., .86., .75., .33., .75., .61., .57

R-INLA code

The following code indicates the workflow and includes the model calls to R-INLA.

```
#Install packages "INLA" from "http://www.r-inla.org/download"
#Get and process data and adjacency matrix formula = Y
-1 + Node.Name + f(Expert.No,model="iid") + f(Node.No,
model="bym2",graph=adj)

result = inla(formula,data=data,family=c("beta","binomial"),
Ntrials = data$n,control.fixed=list(prec.intercept=0.1,prec=0.1),
verbose= FALSE,control.predictor = list(compute=T, link=1))

#Plot the required marginal posterior probabilities

#Fit the model without random effects and compare the IQRs
formula2 = Y   -1 + Node.Name

result2 = inla(formula2,data=data,family=c("beta","binomial"),
Ntrials = dat$n,control.fixed=list(prec.intercept=0.1,prec=0.1),
verbose=FALSE, control.predictor = list(compute=T, link=1))

print((result$summary.fitted.values"0.975quant"[5000:5050] –
result$summary.fitted.values"0.025quant"[5000:5050])/
(result2summary.fitted.values"0.975quant"[5000:5050]-
result2summary.fitted.values"0.025quant"[5000:5050]))

#Plot the different posterior consensus probabilities for the
individual
nodes
```