Freshwater Biology

How good are Bayesian belief networks for environmental management in reality? A test with multiple stressors in a New Zealand river.

Journal:	*Freshwater Biology*
Manuscript ID:	FWB-P-Oct-14-0613.R2
Manuscript Type:	Standard Paper
Date Submitted by the Author:	02-Jul-2015
Complete List of Authors:	Death, Russell; Massey University, Institute of Natural Resources - Ecology Death, Fiona; Aquanet Consulting, STUBBINGTON, RACHEL; Nottingham Trent University, School of Science and Technology Joy, Mike; Massey University, Ecology van den Belt, Marjan
Keywords:	Bioassessment < Applied Issues, Running water / rivers / streams < Habitat, Community < Level of Organisation, Modelling / statistics < Process / Approach / Methods, Invertebrates < Taxonomic Group / Assemblage

How good are Bayesian belief networks for environmental management in reality? A test with data from an agricultural river catchment.

RUSSELL G. DEATH[1], FIONA DEATH[1,2], RACHEL STUBBINGTON[3],

MICHAEL K. JOY[1] AND

MARJAN VAN DEN BELT[4]

[1]*Institute of Agriculture and Environment – Ecology, Massey University, Private Bag 11-222, Palmerston North, New Zealand,* [2]*Aquanet Consulting Ltd. 441 Church St, Palmerston North 4410, New Zealand,* [3]*School of Science and Technology, Nottingham Trent University, Nottingham, Nottinghamshire, NG11 8NS, UK,* [4]*Ecological Economics Research New Zealand, Massey University, Private Bag 11-222, Palmerston North, New Zealand*

Correspondence: Russell Death, Institute of Agriculture and Environment – Ecology, Massey University, Private Bag 11-222, Palmerston North, New Zealand.

Email r.g.death@massey.ac.nz

SUMMARY

1. The ecological health of rivers worldwide continues to decline despite increasing effort and investment in river science and management. Bayesian belief networks (BBNs) are increasingly being used as a mechanism for decision-making in river management because they provide a simple visual framework to explore different management scenarios for the multiple stressors that impact rivers. However, most applications of BBN modelling to resource management use expert knowledge and/or limited real data, and fail to accurately assess the ability of the model to make predictions.
2. We developed a BBN to model ecological condition in a New Zealand river using field/GIS data (from multiple rivers), rather than expert opinion, and assessed its predictive ability on an independent dataset. The developed BBN performed moderately better than a number of other modelling techniques (e.g., artificial neural networks, classification trees, random forest, logistic regression), although model construction was more time-consuming. Thus the predictive ability of BBNs is (in this case at least) on a par with other modelling methods but the approach is distinctly better for its ability to visually present the data linkages, issues and potential outcomes of management options in real time.
3. The BBN suggested management of habitat quality, such as riparian planting, along with the current management focus on limiting nutrient leaching from agricultural land may be most effective in improving ecological condition.
4. BBNs can be a powerful and accurate method of effectively portraying the multiple interacting drivers of environmental condition in an easily understood manner. However, most BBN applications fail to appropriately test the model fit prior to use. We believe this lack of testing may seriously undermine their long-

term effectiveness in resource management, and recommend that BBNs should be used in conjunction with some measure of uncertainty about model predictions. We have demonstrated this for a BBN of ecological condition in a New Zealand river, shown that model fit is better than that for other modelling techniques, and that improving habitat would be equally effective to reducing nutrients to improve ecological condition.

Keywords: Bayesian belief network, decision support, ecological modelling, multiple stressors, predictive ability, resource management, water quality

Introduction

The ecological integrity, biodiversity, water quality and volume of water in rivers and streams are in global decline (Dudgeon *et al.*, 2006; Vorosmarty *et al.*, 2010; Feld *et al.*, 2011). This decline is a result of multiple interacting stressors (Matthaei, Piggott & Townsend, 2010; Wagenhoff *et al.*, 2011; Piggott *et al.*, 2012) including water abstraction for consumptive and agricultural needs (Dewson, James & Death, 2007; Poff & Zimmerman, 2010), invasive species (Olden *et al.*, 2010) channelization, sedimentation, eutrophication (Carpenter *et al.*, 1998; Allan, 2004) and changing climate regimes (Palmer *et al.*, 2008; Death, Fuller & Macklin, in press). The decline in ecological condition is occurring despite unprecedented environmental monitoring (Davies *et al.*, 2010; Friberg *et al.*, 2011), more sophisticated techniques for evaluating collected data (Reynoldson *et al.*, 1997; Boulton, 1999; Linke *et al.*, 2005), increased access to data via frameworks such as online GIS (Snelder & Biggs, 2002; Snelder & Hughey, 2005), more widespread public concern (Cullen, Hughey & Kerr, 2006; Hughey, Kerr & Cullen, 2010) and more ecologically-based legal frameworks (Fore *et al.*, 2008; Acreman & Ferguson, 2010; Hering *et al.*, 2010).

Freshwater management involves assessing the current state and stressors of a waterbody and making educated decisions about the response of that state to changes in the stressors. Although a strong scientific basis underpins many of these decisions, the outcomes from particular options are never certain. However, there is weak, or usually absent, understanding of the relative level of uncertainty associated with alternative options, particularly among resource planners and lawyers associated with the decision-making process (Downes *et al.*, 2002). This uncertainty is further complicated by the typical focus of traditional reductionist science on single stressors, while real-world stressors interact in potentially unknown ways (Downes, 2010;

Ormerod *et al.*, 2010; Harris & Heathwaite, 2012). Effective resource management will require the application of more complex multivariate models that incorporate the stressor interactions and that managers can use to explore different options and their implications (Harris & Heathwaite, 2012) .

Machine learning techniques (e.g., artificial neural networks, classification trees, and random forests) are increasingly being used to describe and predict species' distributions in relation to environmental variables (e.g., Joy & Death, 2004; Elith *et al.*, 2006; Olden, Joy & Death, 2006). Bayesian belief networks (BBNs) are one such graphical, rule-based modelling technique that has emerged as a potentially useful research and management tool (e.g., McCann, Marcot & Ellis, 2006; Uusitalo, 2007; Pourret, Naim & Marcot, 2008). In environmental management BBNs can provide a useful visual depiction of the causal linkages between multiple environmental drivers and ecological state (Aguilera *et al.*, 2011; Allan *et al.*, 2012). They also allow managers to model changes in those drivers to explore the effects on the condition of that ecological state (McCann, Marcot & Ellis, 2006). For example a BBN can be used to investigate how changes in land use may directly and/or indirectly alter invertebrate community composition. BBNs have several advantages: 1) their graphical structure allows easy interpretation by non-modellers (McCann, Marcot & Ellis, 2006); 2) they can be used with incomplete data sets (Uusitalo, 2007); 3) they can incorporate expert knowledge (Pollino *et al.*, 2007; Uusitalo, 2007); 4) they can combine categorical and continuous variables (Marcot *et al.*, 2001); 5) there is an explicitly documented level of uncertainty (Uusitalo, 2007); 6) they can predict in both directions (e.g., water quality can be predicted from the biota present and can also predict what biota will be present with different conditions; (Paisley *et al.*, 2011); and 7) there is relatively inexpensive user-friendly software that allows BBNs to be constructed. One major drawback of available software for BBNs in ecology is arguably the requirement to discretize continuous variables, as most biological data is continuous rather than discrete. Hybrid or non-parametric BBNs offer considerable future potential for the use of continuous variables in BBNs but are not available in commonly used software at present (e.g. Morales-Napoles *et al.*, 2014; Ropero *et al.*, 2014).

Reflecting these advantages, BBNs have increasingly been applied to model environmental outcomes (Pourret, Naim & Marcot, 2008; Aguilera *et al.*, 2011). Published accounts of the application of these models have stressed the ability of BBNs to integrate expert knowledge (Pollino *et al.*, 2007; Aguilera *et al.*, 2011). They are often used simply as a heuristic framework for identifying the important issues, environmental drivers and potential interactions that require consideration. In a review of BBN applications in environmental modelling, Aguilera et al. (2011) found that a third of studies used only expert opinion to build the BBN, while over half had no independent assessment of the accuracy of the BBN predictions. While experts have considerable knowledge, their ability to integrate that knowledge in an objective multivariate and predictive manner can be limited, particularly if there is no independent validation of their assessment (Marcot, 2012). Effective use of BBNs in resource management must rest not only on their ability to integrate differing forms of data and visually depict potential linkages, but also on their ability to make accurate predictions. More frequent and confident BBN use in resource management requires greater emphasis on constructing BBNs using data and/or independently assessing model fit against that data (Marcot, 2012).

We used field-collected and GIS data to develop a BBN model for an invertebrate-based measure of ecological condition (QMCI; Quantitative Macroinvertebrate

Community Index (Stark, 1985)) in the Manawatu River, New Zealand. We then used the BBN to explore management options for improving the QMCI in river reaches where it is currently low. We also used the BBN to predict ecological condition in unsampled river reaches and to map the results in a GIS. Before using the BBN to examine different management scenarios, we independently (of the data used to build the model) assessed its predictive ability and compared that to the accuracy of other linear and machine-learning approaches. We hypothesised that the BBN would outperform the linear model, but would perform similarly to the other machine-learning techniques in its ability to model ecological condition.

Methods

Study area

The Manawatu River (catchment area 3694 km^2) is a 7th order river in the southern North Island of New Zealand. It arises in native forest in the Tararua and Ruahine Mountain Ranges then flows through predominantly sheep, beef and dairy farmland for most of its length. This land use, along with variably treated sewage discharges from several small towns and one city (Palmerston North, population 78,800) has contributed to the Manawatu River having some of the highest nutrient levels, gross primary production, deposited sediment and lowest water quality in New Zealand (Roygard, McArthur & Clark, 2012). In response to public concern over the river's condition, NZ$30 m was allocated in 2011 by local and regional government to ameliorate some stressors by stream fencing, riparian planting, sewage plant upgrades and nitrate leaching restrictions.

Data from over 40 river catchments in the southern North Island (Fig. 1) were also used to help with BBN construction (see below). This larger area generally has similar characteristics to the Manawatu River catchment, with rivers and streams arising in higher-altitude forested areas before flowing downstream through predominantly pastoral farmland.

Data sets used in model development

Invertebrate communities were sampled at 963 sites throughout the southern North Island (including 194 sites in the Manawatu River catchment; Fig. 1), during studies conducted at Massey University between 1994 and 2007. Most of these sampling occasions involved 5 replicate 0.1 m^2 Surber samples, although some collections comprised a single 1-minute kick-net sample (see Death & Joy, (2004) for more details). Samples were filtered through a 500 µm mesh sieve and identified to the lowest possible taxonomic level (usually genera) using Winterbourn, Gregson & Dolphin (2006). Where samples were collected from a site in multiple years, only the most recent was used in the analysis. The QMCI was used as an index of biological water quality that incorporates the pollution sensitivity and abundance of genera. It is relatively independent of sampling effort and season (Duggan, Collier & Lambert, 2002), and we are therefore confident that the measures of biological water quality used are an accurate representation of ecological condition, even though data were collected for a variety of reasons.

Eighty-five catchment and reach GIS variables (Appendix 1) were extracted for each site from the River Environment Classification (REC; Snelder & Biggs (2002) or Freshwater Ecosystems of New Zealand (FENZ) geodatabase (Leathwick *et al.*, 2010). These variables included environmental measures likely to influence instream biology at the reach and/or catchment level, such as land cover, land use pressures, typography, geology and climate (for more detail see Snelder, Dey & Leathwick (2005) and Wild *et al.*,(2005)). The variables were derived for each section of the region's river network (average length = 700 m) by modelling variables from a 30 m Digital Elevation Model and/or digitised 1:50,000 maps of typography, geological rock type or land cover from the REC (Snelder & Biggs, 2002). All variables had been weighted by total annual runoff; however, for the study streams, these were highly correlated with the unweighted variables and were thus excluded from further analysis.

Variable selection and discretisation

QMCI is an index of ecological condition used for invertebrate bioassessment in New Zealand rivers that combines the abundance of taxa in a sample and a sensitivity score for that taxa (from 1 – 10). Higher QMCI indicates better ecological condition and range from 0 to 10. QMCI scores are continuous but are usually assigned to one of four classes for assessing water quality (Boothroyd & Stark, 2000). However, as there were few sample sites in the two intermediate QMCI categories, these were pooled into one moderate category for analysis. Thus, in the full data set, 496 (51%) sites were classed as clean (QMCI ≥ 6), 262 (27%) as moderate (QMCI $4 > x > 6$) and 205 (21%) as poor (QMCI ≤ 4) ecological condition. Of these data, 194 sites were in the Manawatu River catchment, where there were 105 (54%) sites in the clean, 49 (26%) sites in the moderate and 40 (21%) sites in the poor categories.

For the 85 GIS variables for potential inclusion in the BBN, strongly correlated variables ($r > 0.8$) were represented by only one of the correlates leaving 47 potentials for inclusion in the BBN. The Waikato Environment for Knowledge Analysis

(WEKA) machine learning software (version 3.6.1) (Witten, Frank & Hall, 2011) was used to further reduce the dataset using the CfsSubsetEval (correlation-based feature subset selection) procedure and BestFirst selection method that evaluates the individual predictive ability of each variable along with the degree of redundancy (Witten, 1999). BestFirst and CfsSubsetEval are both options in the WEKA attribute selection procedure. BestFirst is a mutual information search method that searches the attribute subset space by greedy hillclimbing augmented with a backtracking facility (Witten, Frank & Hall, 2011). CfsSubsetEval evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature to predict, in this case QMCI, along with the degree of redundancy between the subsets (Hall, 1998; Witten, Frank & Hall, 2011). This reduced the list of potential variables to nine, comprising four measures of water chemistry (nitrogen, phosphorus, calcium and hardness), upstream slope, one measure of catchment land use (percent pasture) and three measures of reach habitat condition (riparian shade, native riparian vegetation and habitat type). Habitat type (ReachHab) is a weighted average of proportional cover of flow types (1–still; 2–backwater; 3–pool; 4–run; 5–riffle; 6–rapid; 7–cascade, predicted from a boosted regression tree model using GIS variables and is detailed in Leathwick *et al.* (2008).

Bayesian Belief Network (BBN) construction

The BBN was constructed using Netica^{TM} 5.02 (Pourret, Naim & Marcot, 2008). The network of interconnected variables is represented as a series of nodes (Fig 2). Each node has potential states for that variable e.g., good, moderate or poor. To discretise continuous GIS variables into states within a node, classification tree analysis (De'ath & Fabricius, 2000) was used to model QMCI group membership from the nine predictor variables using WEKA. Thresholds for node states were garnered from the associated numerical values at any branch in the classification tree where that variable was important (Table 1). Thus the number of states differs across variables based on the number of branches involving that variable in the classification tree. Causation flows from a 'parent' node to a 'child' node.

Two intermediate nodes in the network, one for water chemistry (Waterchem) and one for habitat quality (HabitatQual), were added (Fig. 2) to link their respective parent nodes and reduce the number of inputs into the QMCI node (Marcot et al. 2006). Conditional Probability Tables (CPTs) were developed with the expectation-maximization algorithm (EM Learning) in Netica™ from the compiled data. The expectation–maximization (EM) algorithm is an iterative method for finding maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables (Do & Batzoglou, 2008). CPTs calculate the probability of each state in a node occurring, given each combination of conditions in the parent (input) nodes (Pourret, Naim & Marcot, 2008).

Alternative classification performance

To compare the predictive ability of the BBN to other classification methods, a range of linear (i.e., logistic regression) and machine-learning (i.e., classification trees, random forests and artificial neural networks) techniques were applied to the same data as the BBN (i.e., predicting three QMCI classes with nine variables). Logistic regression was conducted with a multinomial logistic regression model and ridge estimator; a modified form of the technique of le Cessie & van Houwelingen (1992) (Witten & Frank, 2000). A classification tree was generated with the j48 option in

WEKA that generates a pruned or unpruned C4.5 decision tree (Quinlan, 1993; Witten & Frank, 2000). Artificial neural networks used backpropogation to classify with a learning rate of 0.3 and 6 hidden layers (Witten & Frank, 2000). The same nine GIS variables were used for these models, but in their quantitative form. WEKA was used for all four analyses to model QMCI from those GIS variables.

Models were evaluated by hold-out validation with a randomly selected 10% subset of the training data. There are a wide range of metrics that can be used to evaluate model fit and performance (for a detailed review see Witten, Frank & Hall (2011) and Marcot (2012)). We used several commonly used metrics that assess both raw predictive ability and ability relative to occurrence. The percentage of incorrect predictions (percent error) is a simple, easily understood metric but is sensitive to the number and size of the nodes. For example, if you have a very common state in the node and predict it will always occur (P=1.0) then you have a high probability of being correct simply because it usually occurs. Area under receiver operating characteristic curves (ROC) attempt to correct for this by plotting true positives against false positives to search for a balance between sensitivity and specificity (Hand, 1997). They range from 1 to 0, with 0.5 denoting totally random models and >0.5 improvement on random (Marcot, 2012). For BBNs Spherical payoff is similar to the area under receiver operating curves (Hand, 1997; Marcot, 2012). Cohen's kappa also ranges from 0 to 1, with 1 being perfect classification that also assesses correct predictions relative to how common a state actually is (Boyce *et al.*, 2002; Olden, Lawler & Poff, 2008). The logarithmic loss score (Dlamini, 2010) was used to compare BBNs of alternate architecture. The index ranges from 0 to infinity, with 0 the best possible score. Unlike the indices above that must be calculated outside

Netica™, this index is provided within the program and gives a quick metric for evaluating alternate BBNs.

Making predictions of current state from the BBN

The developed BBN was used to predict the QMCI state for all river reaches in the Manawatu River catchment using the available GIS data sets and the results plotted in ESRI ArcMap™ 9.3.1. To determine the probability level at which a site was designated as poor, moderate or clean, the training data set (without the QMCI data) was run through the BBN. The output QMCI state predictions were then entered into Schröder's ROC Plotting and AUC Calculation Transferability Test software (Schröder & Richter, 2000) to determine the critical probability to indicate a particular QMCI state. The Area Under the ROC-Curve (AUC) is a threshold-independent measure of predictive performance with bootstrapped confidence intervals calculated using the percentile method (Buckland, Burnham & Augustin, 1997; Augustin, Mugglestone & Buckland, 1998). It seeks to optimise cut-off probabilities to indicate a particular class with respect to i) maximised Kappa, ii) minimised difference between sensitivity and specificity and iii) maximised correct classification rate taking into account different costs of false positive or false negative predictions. For example for a rare state a probability of 0.3, rather than say 0.5, might provide the greatest likelihood of a correct prediction, but at the same time the least likelihood of a false positive. Thus if the model predicts this particular class with a probability greater than 0.3 (the critical probability from the AUC) then we allocate that class (e.g. clean QMCI) as the prediction of the model. Thus a clean state was designated for a site if the probability for a clean state was greater than or equal to 0.473, $P \geq 0.625$ for a moderate state and $P \geq 0.523$ for a poor state.

Results

Bayesian Belief Network evaluation

The architecture of the BBN for predicting QMCI state is presented in Figure 2. The ability of the network to describe the training data (the Manawatu catchment sites, n=194) was good. There was a 17.5% error rate, a logarithmic loss score of 0.38 (this ranges from 0 to infinity, with 0 the best possible score), a spherical payoff score of 0.87 (this ranges from 0 to 1, with 1 being the best possible score) and a Cohen's kappa of 0.70 (indicative of good model fit) (Landis & Koch, 1977). However, the BBN did not perform as well on an independent data set (i.e., regional sites not in the Manawatu River catchment, n=769): there was a 46.8% error rate, a log loss score of 3.09, a spherical payoff score of 0.57 and a Cohen's kappa of 0.21. A BBN built from the independent non-Manawatu test data (n=769) and evaluated against the Manawatu training data (n=194) performed better than the BBN model built on the Manawatu data: the error rate was 36.1%, log loss score 1.14, spherical payoff 0.71 and Cohen's kappa 0.41.

The final BBN was built with 662 randomly selected sites from the full 963 site data set and was independently tested with the remaining held out 300 sites. This was repeated 10 times and the best performing model retained. The model error rate on the independent data set was 28.0% error rate, log loss score of 0.72, spherical payoff of 0.76 and a Cohen's kappa of 0.53 (indicative of moderate model fit).

Alternative modelling evaluation

The ability of the four alternative classification techniques to describe the training data and predict test data (i.e. 10% of the training data) were slightly lower than those of the best BBN (Table 2). Error rates were between 33 and 38% and Cohen's kappa values were between 0.39 and 0.44. There was not really much difference between the approaches used but the simple logistic regression performed best with respect to percent error and Cohen's kappa, but Random forests with respect to AUC (Table 2).

Mapping the predictions

The QMCI predictions from the BBN for the Manawatu catchment are mapped in Figure 3. The good ecological condition of streams draining the native vegetation of the central Tararua and Ruahine State Forest Park is clear. Further downstream, the cumulative agricultural effects shift the streams from moderate to poor in the western Manawatu Plains and in the high-density dairy farmland on the eastern side of the Ruahine Ranges around the town of Dannevirke. Large areas of the south-eastern catchment have waterways with moderate ecological condition reflecting the lower intensity agricultural land use in this region.

Model predictions

To investigate management options for improving ecological condition within the BBN, the effects of changes to water chemistry, habitat quality and the amount of catchment in pasture were examined (Table 3). Not surprisingly, altering the state (i.e., good, moderate, poor) of the variables directly linked to the QMCI node (WaterChem and HabitatQual; Fig. 2) had the biggest effect on QMCI condition.

Changing habitat condition from good to poor resulted in a reduction in the likelihood of a clean site by 0.53 and an increase in the chance of a poor site by 0.29. Changes to the water chemistry node (WaterChem) from good to poor had a similar effect in increasing the chance of a poor site by 0.26 and decreasing the chance of a clean site by 0.58 (Table 3). The state of the Habitat Quality node was determined by the percent of the catchment in pasture (USPasture), riparian vegetation (SegRipShad) and ReachHab (type of microhabitats). The state of the Water Chemistry node was determined by water hardness and concentrations of nitrogen, phosphorous, and calcium in the upstream reaches. Thus reduction in pasture catchment land use and/or increased riparian planting to increase shade yielded, or reductions in inflowing nutrients had similar effects on QMCI (Table 3). If all nodes were held constant except USPasture and its' linked nodes, the biggest change in QMCI state (a reduction of 0.19 in the likelihood of a clean stream) occurred with a shift from low to moderate state (i.e., as percent catchment in pasture increased above 28%; Tables 1 and 3). There were minimal changes (i.e. reductions in chance of a clean stream of 0.02 or 0.05) to QMCI between the very high (>90%), high (>68%) and moderate (>28%) pasture states.

Altering the state of other nodes not directly linked to the QMCI node (e.g. USCalcium, USHardness LogNConcen; Fig. 2) had considerably less effect on the state of the QMCI node in the BBN (results not presented). For example, changing the state of the Nitrogen node (LogNConcen) only had a weak effect on the QMCI value (increasing the LogNConcen from below 0.1 mg/l (low) to above (high) resulted in a 14% lower chance of getting a clean stream and a 6% increase in the chance of a poor stream) . However, for this all unconnected nodes were left fixed, and it is questionable how often this would actually occur in the real world. It seems unlikely streams with high nitrogen levels actually occur in the Manawatu catchment that are not also high in phosphorus, percent catchment in pasture and low in riparian shade.

Discussion

Bayesian Belief Network performance

Although Bayesian Belief Networks have been advocated by many as effective environmental management tools there have been few efforts to evaluate their predictive ability and/or performance relative to other modelling techniques (Uusitalo, 2007; Aguilera *et al.*, 2011; Allan *et al.*, 2012; Marcot, 2012). We constructed a BBN of ecological condition in a New Zealand river using data, rather than expert opinion, and evaluated the network predictive ability on independent data. The BBN performed moderately well when evaluated with the independent data and we would consequently be confident that management decisions made with the network are accurate. Few previous studies that have evaluated BBNs with independent data used the same evaluation metrics (which makes comparison difficult) for model fit that we applied, but the BBN used by Marcot (2012) to predict the age of martens had spherical payoff indices between 0.70 and 0.96, similar to ours. If BBNs are to become credible tools for integrating data and making resource management decisions their predictive ability must be evaluated appropriately. No one would consider presenting a regression equation for use in management unless it was statistically significant and with an appropriately high r^2, yet many BBNs are presented without any measure of how good the network is (Aguilera *et al.*, 2011).

Marcot (2012) provides a comprehensive review of metrics that can be used to evaluate different aspects of BBN performance and uncertainty. However, as a balance between statistical rigour and ease of calculation, we recommend, as a minimum, to report the error rate, log loss score, spherical payoff score and Cohen's kappa for the training data set used to build the BBN. Furthermore, if the BBN is to be used to make predictions or evaluate management alternatives, then these same statistics for an independent test data set should also be reported. This will be more challenging for BBNs that are developed with expert knowledge; however we still believe that some form of independent assessment of the accuracy of model predictions are needed before the BBN is used. This might involve consulting experts not involved in the original workshops, or comparing outcomes with some real examples of the possible combinations in the BBN.

We are not aware of any previous evaluations of whether BBNs perform better at modelling than other classification techniques. For our data on ecological condition in a New Zealand river, the BBN did perform moderately better than all other techniques (i.e., logistic regression, classification trees, random forests, artificial neural networks). These other classification techniques may be easier and quicker to use than a BBN because data does not require discretization. To hasten BBN development, the BBN software can perform discretization on entered quantitative data, often into evenly sized or distributed data chunks; however, the outcomes often do not yield accurate predictions. Most biological data does not fit nicely into discrete groups, and such discretization strategies are thus fraught with problems. We avoided such problems by using classification trees to identify thresholds for the node classes that provided the most informative links with QMCI. BBN accuracy indicated that this worked well, although we did not directly compare the approach with other discretization strategies such as evenly divided groups. In the future Hybrid or non-parametric BBNs, where discretization is not needed, may avoid the need for this approach in BBN construction (e.g. Morales-Napoles *et al.*, 2014; Ropero *et al.*, 2014).

For communicating research findings to the resource managers who use a developed model, the visual linkage diagram (e.g. Fig. 2) and the ability to observe the effect of changing variables immediately with a simple click of the mouse (by clicking on different node states in Netica™) are major advantages of the BBN over the other modelling approaches. Compared to other classification techniques, it also offers the unique benefit of being able to both predict biota from the environment and/or diagnose environmental condition from the taxa present (Paisley *et al.*, 2011). Therefore if the predictive ability of a BBN is good (as assessed on independent data) then the BBN will be a superior technique for communicating the complex interaction of multiple stressors and for evaluating the efficacy of alternative management options.

Managing ecological condition with the Bayesian Belief Network

The constructed BBN provided a good description of the combined outcome of the multiple stressors affecting one measure of ecological condition (QMCI) in the Manawatu River. This is reassuring given that all the environmental variables used were GIS-derived, rather than actual field measures; however, they were still able to accurately predict QMCI values across the catchment. We had chosen GIS variables specifically so we could also predict ecological condition in reaches where invertebrates had not been sampled. It is then possible to map the predicted patterns at a catchment level for a more complete perspective on ecological condition. The map facilitated observation of the gradual change in ecological condition from the pristine mountain streams in the forest parks to increasing degradation as the river flows downstream through increasingly agricultural land use at lower elevations. Similarly, the map clearly shows the more dramatic transition from clean to poor ecological condition in streams and rivers around the town of Dannevirke, where dairy farming is more intensive. Linking the BBN directly with a GIS map would further enhance the ability of the BBN to communicate outcomes from alternative management decisions.

The BBN provides some interesting insight into the potential options for improving ecological condition (as measured by QMCI) in the Manawatu River. For example, a current focus of several environmental management and advocacy groups involved in enabling the river clean-up fund is to limit nutrient leaching into the river from agricultural land (Roygard, McArthur & Clark, 2012). However, our model suggests that modification of the stream and riparian habitat may be equally effective to nutrient management in increasing the QMCI, despite the strong links between agriculture, nutrient levels and the QMCI found elsewhere (Boothroyd & Stark, 2000; Wagenhoff *et al.*, 2011; Clapcott *et al.*, 2012). Increases in QMCI may come from improving habitat quality through provision of more riparian shading or reducing nutrient leaching from the land. Furthermore, we observed a threshold (28%) for the percentage of pastoral agriculture in the catchment that will degrade ecological condition, with further increases in agricultural land use having few concomitant effects on condition, a phenomenon also found elsewhere in New Zealand (Death & Collier, 2010) (but see Wagenhoff *et al.*, 2011).

When evaluating management options, it is important to avoid presenting unrealistic scenarios to the BBN. For example, changing the nitrogen node status had a limited effect on QMCI, despite the known links between nitrogen and QMCI (Clapcott *et al.*, 2012). This partly reflects the fact that many node combinations are unlikely to occur in reality, such as high levels of upstream pastoral agriculture and low levels of nutrients; such combinations therefore remain untested by the model. However, in many cases these are the scenarios that require evaluation, for example to determine if reducing nitrogen leaching in pastoral catchments can improve ecological condition. This is one disadvantage of models constructed from real world data, rather than expert opinion; although the latter must also be based on limited real world experience. If such scenarios are explored it therefore needs to be stressed to users the difficulty of modelling artificial scenarios where variables are manipulated outside those normally seen in reality, e.g., can we really ever have a catchment with a high percentage of pasture and low nutrient levels? That said, it may still be useful to explore the effect of increasing percentage of pastoral catchment in the BBN model, while keeping other factors constant, and observe that the biggest effect on the QMCI is the change from little or no agriculture (< 28% pasture) to some, with no dramatic decreases in ecological condition as percent pasture in the upstream catchment gets to be very high.

Conclusion

Despite the caveats above, the BBN is a very useful tool for exploring the efficacy of different management options before expending time and money in activities that may not work, and/or for which there will be a considerable time lag before outcomes are clear (Parkyn *et al.*, 2003). It is also very easy for non-experts to select different node states and explore the ecological outcomes once the model is built, without any need to conduct complicated analysis (c.f., artificial neural networks); provided they appreciate some node combinations may not occur in reality. Therefore, provided that the BBN makes accurate predictions they are considerably more useful, than other classification techniques for modelling, forecasting and management of resources exposed to multiple interacting stressors.

The BBN constructed to predict ecological condition in the Manawatu River catchment performed well on independent data. It provided some insightful direction as to the best strategies for improving ecological condition of the river, which have not been given much attention prior to this. The BBN also allowed mapping of catchment wide ecological condition, which in turn provided a more holistic perspective on the water management issues. We support the view of many others (e.g. McCann, Marcot & Ellis, 2006; Pollino *et al.*, 2007; Uusitalo, 2007; Aguilera *et al.*, 2011; Leigh *et al.*, 2011; Allan *et al.*, 2012) that BBNs are extremely useful and intuitive tools for understanding and managing environmental issues. Although modelling of any kind can seem difficult for non-experts; the readily available software makes the task relatively straightforward. We have outlined the general steps in BBN development in Table 4, many of which do not require specialised software. The most important step is the appropriate testing and reporting of their efficacy (Table 4). If the developed models are not independently validated, but are still used for management decisions because they appear rigorous to non-scientist managers, the outcomes are likely to be different from those expected and undermine the power BBN modelling could offer environmental management.

Acknowledgements

Thank you to Vicky Forgie for her pleasant humour in the support of this project. Thanks also to Regina Magierowski for comments on an early draft of the manuscript. The authors thank the Ministry for Science and Innovation for funding the Integrated Freshwater Solutions project (Contract MAUX1002).

References

Acreman M.C. & Ferguson A.J.D. (2010) Environmental flows and the European Water Framework Directive. *Freshwater Biology,* **55,** 32-48.

Aguilera P.A., Fernandez A., Fernandez R., Rumi R. & Salmeron A. (2011) Bayesian networks in environmental modelling. *Environmental Modelling & Software,* **26,** 1376-1388.

Allan J.D. (2004) Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology Evolution and Systematics,* **35,** 257-284.

Allan J.D., Yuan L.L., Black P., Stockton T.O.M., Davies P.E., Magierowski R.H. & Read S.M. (2012) Investigating the relationships between environmental stressors and stream condition using Bayesian belief networks. *Freshwater Biology,* **57,** 58-73.

Augustin N.H., Mugglestone M.A. & Buckland S.T. (1998) The role of simulation in modelling spatially correlated data. *Environmetrics,* **9,** 175-196.

Boothroyd I.K.G. & Stark J.D. (2000) Use of invertebrates in monitoring. In: *New Zealand Stream Invertebrates: Ecology and Implications for Management.* (Eds K.J. Collier & M.J. Winterbourn), pp. 344-373. New Zealand Limnological Society, Hamilton.

Boulton A.J. (1999) An overview of river health assessment: philosophies, practice, problems and prognosis. *Freshwater Biology,* **41,** 469-479.

Boyce M.S., Vernier P.R., Nielsen S.E. & Schmiegelow F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling,* **157,** 281-300.

Buckland S.T., Burnham K.P. & Augustin N.H. (1997) Model selection: An integral part of inference. *Biometrics,* **53,** 603-618.

Carpenter S.R., Caraco N.F., Correll D.L., Howarth R.W., Sharpley A.N. & Smith V.H. (1998) Nonpoint pollution of surface waters with phosphorus and nitrogen. In: *Issues in Ecology No. 3.* Ecological Society of America.

Clapcott J.E., Collier K.J., Death R.G., Goodwin E.O., Harding J.S., Kelly D., Leathwick J.R. & Young R.G. (2012) Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshwater Biology,* **57,** 74-90.

Cullen R., Hughey K. & Kerr G. (2006) New Zealand freshwater management and agricultural impacts. *Australian Journal of Agricultural and Resource Economics,* **50,** 327-346.

Davies P.E., Harris J.H., Hillman T.J. & Walker K.F. (2010) The Sustainable Rivers Audit: assessing river ecosystem health in the Murray-Darling Basin, Australia. *Marine and Freshwater Research,* **61,** 764-777.

De'ath G. & Fabricius K.E. (2000) Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology,* **81,** 3178-3192.

Death R.G. & Collier K.J. (2010) Measuring stream macroinvertebrate responses to gradients of vegetation cover: when is enough enough? *Freshwater Biology,* **55,** 1447-1464.

Death R.G., Fuller I.C. & Macklin M.G. (in press) Resetting the river template: the potential for climate-related extreme floods to transform river geomorphology and ecology. *Freshwater Biology*.

Death R.G. & Joy M.K. (2004) Invertebrate community structure in streams of the Manawatu-Wanganui region, New Zealand: the roles of catchment versus reach scale influences. *Freshwater Biology,* **49,** 982-997.

Dewson Z.S., James A.B.W. & Death R.G. (2007) A review of the consequences of decreased flow for instream habitat and macroinvertebrates. *Journal of the North American Benthological Society,* **26,** 401-415.

Dlamini W.M. (2010) A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling & Software,* **25,** 199-208.

Do C.B. & Batzoglou S. (2008) What is the expectation maximization algorithm? *Nat Biotech,* **26,** 897-899.

Downes B.J. (2010) Back to the future: little-used tools and principles of scientific inference can help disentangle effects of multiple stressors on freshwater ecosystems. *Freshwater Biology,* **55,** 60-79.

Downes B.J., Barmuta L.A., Fairweather P.G., Faith D.P., Keough M.J., Lake P.S., Mapstone B.D. & Quinn G.P. (2002) *Monitoring ecological impacts : concepts and practice in flowing waters,* Cambridge University Press, Cambridge.

Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.I., Knowler D.J., Leveque C., Naiman R.J., Prieur-Richard A.H., Soto D., Stiassny M.L.J. & Sullivan C.A. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews,* **81,** 163-182.

Duggan I.C., Collier K.J. & Lambert P.W. (2002) Evaluation of invertebrate biometrics and the influence of subsample size, using data from some Westland, New Zealand, lowland streams. *New Zealand Journal of Marine and Freshwater Research,* **36,** 117-128.

Elith J., Graham C., H., Anderson R., P., Dudik M., Ferrier S., Guisan A., Hijmans R., J., Huettmann F., Leathwick J., R., Lehmann A., Li J., Lohmann L., G., Loiselle B., A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J., M., Townsend P., A., Phillips S., J., Richardson K., Scachetti-Pereira R., Schapire R., E., Soberon J., Williams S., Wisz M., S. & Zimmermann N., E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography,* **29,** 129-151.

Feld C.K., Birk S., Bradley D.C., Hering D., Kail J., Marzin A., Melcher A., Nemitz D., Pedersen M.L., Pletterbauer F., Pont D., Verdonschot P.F.M. & Friberg N. (2011) From natural to degraded rivers and back again: a test of restoration ecology theory and practice. In: *Advances in Ecological Research, Vol 44.* (Ed G. Woodward), pp. 119-209. Advances in Ecological Research. Elsevier Academic Press Inc, San Diego.

Fore L.S., Karr J.R., Fisher W.S. & Davis W.S. (2008) Making Waves with the Clean Water Act. *Science,* **322,** 1788-1788.

Friberg N., Bonada N., Bradley D.C., Dunbar M.J., Edwards F.K., Grey J., Hayes R.B., Hildrew A.G., Lamouroux N., Trimmer M. & Woodward G. (2011) Biomonitoring of Human Impacts in Freshwater Ecosystems: The Good, the Bad and the Ugly. In: *Advances in Ecological Research, Vol 44.* (Ed G. Woodward), pp. 1-68. Advances in Ecological Research. Elsevier Academic Press Inc, San Diego.

Hall M.A. (1998) Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand.

Hand D.J. (1997) *Construction and assessment of classification rules,* Wiley, Chichester ; New York.

Harris G.P. & Heathwaite A.L. (2012) Why is achieving good ecological outcomes in rivers so difficult? *Freshwater Biology,* **57,** 91-107.

Hering D., Borja A., Carstensen J., Carvalho L., Elliott M., Feld C.K., Heiskanen A.S., Johnson R.K., Moe J., Pont D., Solheim A.L. & Van De Bund W. (2010) The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. *Science of the Total Environment,* **408,** 4007-4019.

Hughey K.F.D., Kerr G.N. & Cullen R. (2010) Public perceptions of New Zealand's environment: 2010. Lincoln University, Lincoln.

Joy M.K. & Death R.G. (2004) Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshwater Biology,* **49,** 1036-1052.

Landis J.R. & Koch G.G. (1977) Measurement of observer agreement for categorical data. *Biometrics,* **33,** 159-174.

Le Cessie S. & Van Houwelingen J.C. (1992) Ridge Estimators in Logistic Regression. *Applied Statistics.,* **41,** 191-201.

Leathwick J., West D., Gerbeaux P., Kelly D., Robertson H., Brown D., Chadderton W.L. & Ausseil A.-G. (2010) Freshwater Ecosystems of New Zealand (FENZ) Geodatabase. Users Guide. Department of Conservation, Wellington.

Leathwick J.R., Elith J., Chadderton W.L., Rowe D. & Hastie T. (2008) Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography,* **35,** 1481-1497.

Leigh C., Stewart-Koster B., Sheldon F. & Burford M.A. (2011) Understanding multiple ecological responses to anthropogenic disturbance: rivers and potential flow regime change. *Ecological Applications,* **22,** 250-263.

Linke S., Norris R.H., Faith D.P. & Stockwell D. (2005) ANNA: A new prediction method for bioassessment programs. *Freshwater Biology,* **50,** 147-158.

Marcot B.G. (2012) Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling,* **230,** 50-62.

Marcot B.G., Holthausen R.S., Raphael M.G., Rowland M.M. & Wisdom M.J. (2001) Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management,* **153,** 29-42.

Matthaei C.D., Piggott J.J. & Townsend C.R. (2010) Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology,* **47,** 639-649.

Mccann R.K., Marcot B.G. & Ellis R. (2006) Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **36,** 3053-3062.

Morales-Napoles O., Delgado-Hernandez D.J., De-Leon-Escobedo D. & Arteaga-Arcos J.C. (2014) A continuous Bayesian network for earth dams' risk assessment: methodology and quantification. *Structure and Infrastructure Engineering,* **10,** 589-603.

Olden J.D., Joy M.K. & Death R.G. (2006) Rediscovering the species in community-wide predictive modelling. *Ecological Applications,* **16,** 1449-1460.

Olden J.D., Kennard M.J., Leprieur F., Tedesco P.A., Winemiller K.O. & García-Berthou E. (2010) Conservation biogeography of freshwater fishes: recent progress and future challenges. *Diversity and Distributions,* **16,** 496-513.

Olden J.D., Lawler J.J. & Poff N.L. (2008) Machine Learning Methods Without Tears: A Primer for Ecologists. *The Quarterly Review of Biology,* **83,** 171-193.

Ormerod S.J., Dobson M., Hildrew A.G. & Townsend C.R. (2010) Multiple stressors in freshwater ecosystems. *Freshwater Biology,* **55,** 1-4.

Paisley M.F., Trigg D.J., Martin R., Walley W.J., Andriaenssens V., Buxton R. & O'connor M. (2011) Refinement of artificial intelligence-based systems for diagnosing and predicting river health. p. 259, Vol. SC030189. Environment Agency, Bristol.

Palmer M.A., Reidy Liermann C.A., Nilsson C., Floumlrke M., Alcamo J., Lake P.S. & Bond N. (2008) Climate change and the world's river basins: anticipating management options. *Frontiers in Ecology and the Environment***,** 81-89.

Parkyn S.M., Davies-Colley R.J., Halliday N.J., Costley K.J. & Croker G.F. (2003) Planted Riparian Buffer Zones in New Zealand: Do They Live Up to Expectations? *Restoration Ecology,* **11,** 436-447.

Piggott J.J., Lange K., Townsend C.R. & Matthaei C.D. (2012) Multiple Stressors in Agricultural Streams: A Mesocosm Study of Interactions among Raised Water Temperature, Sediment Addition and Nutrient Enrichment. *Plos One,* **7**.

Poff N.L. & Zimmerman J.K.H. (2010) Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology,* **55,** 194-205.

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

Pourret O., Naim P. & Marcot B. (2008) Bayesian Networks: A Practical Guide to Applications. John Wiley & Sons, Ltd., Chichester. England.

Quinlan. (1993) C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

Reynoldson T.B., Norris R.H., Resh V.H., Day K.E. & Rosenberg D.M. (1997) The reference condition: a comparison of multimetric and multivariate appraoches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society,* **16,** 833-852.

Ropero R.F., Aguilera P.A., Fernandez A. & Rumi R. (2014) Regression using hybrid Bayesian networks: Modelling landscape-socioeconomy relationships. *Environmental Modelling & Software,* **57,** 127-137.

Roygard J.K.F., Mcarthur K.J. & Clark M.E. (2012) Diffuse contributions dominate over point sources of soluble nutrients in two sub-catchments of the Manawatu River, New Zealand. *New Zealand Journal of Marine and Freshwater Research,* **46,** 219-241.

Schröder B. & Richter O. (2000) Are habitat models transferable in space and time? *Journal of Nature Conservation,* **8,** 195-205.

Snelder T., Dey K. & Leathwick J. (2005) Definition of the multivariate environmental river classification - “Freshwater Environments of New Zealand”. Vol. NIWA Client Report: CHC2005-049. NIWA, Christchurch.

Snelder T. & Guest P. (2000) The 'river ecosystem management framework' and the use of river environment classification as a tool for planning. NIWA Client Report: CHC00/81., Christchurch, New Zealand.

Snelder T.B., B. J. F.; Shankar, U.; Mcdowall, B.; Stephens, T.; Boothroyd, I. K. G. (1998) Development of a system of physically based habitat classification for water resources management of New Zealand rivers. NIWA Client report: CHC98/66, Christchurch, New Zealand.

Snelder T.H. & Biggs B.J.F. (2002) Multiscale River Environment Classification for water resources management. *Journal of the American Water Resources Association,* **38,** 1225-1239.

Snelder T.H., Cattaneo F., Suren A.M. & Biggs B.J.E. (2004) Is the River Environment Classification an improved landscape-scale classification of rivers? *Journal of the North American Benthological Society,* **23,** 580-598.

Snelder T.H. & Hughey K.F.D. (2005) The use of an ecologic classification to improve water resource planning in New Zealand. *Environmental Management,* **36,** 741-756.

Stark J.D. (1985) A macroinvertebrate community index of water quality for stony streams., p. 52. Ministry of Works and Development, Wellington.

Uusitalo L. (2007) Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling,* **203,** 312-318.

Vorosmarty C.J., Mcintyre P.B., Gessner M.O., Dudgeon D., Prusevich A., Green P., Glidden S., Bunn S.E., Sullivan C.A., Liermann C.R. & Davies P.M. (2010) Global threats to human water security and river biodiversity. *Nature,* **467,** 555-561.

Wagenhoff A., Townsend C.R., Phillips N. & Matthaei C.D. (2011) Subsidy-stress and multiple-stressor effects along gradients of deposited fine sediment and dissolved nutrients in a regional set of streams and rivers. *Freshwater Biology,* **56,** 1916-1936.

Wild M., Snelder T., Leathwick D.M., Shankar U. & Hurren H. (2005) Environmental variables for the Freshwater Environments of New Zealand classification. NIWAQ client report CHC2004-086., Hamilton.

Winterbourn M.J., Gregson K.L.D. & Dolphin C.H. (2006) *Guide to the aquatic insects of New Zealand. Fourth edition.*

Witten I.H. (1999) *WEKA : practical machine learning tools and techniques with Java implementations,* Dept. of Computer Science University of Waikato, Hamilton, N.Z.

Witten I.H. & Frank E. (2000) *WEKA Machine Learning Algorithms in Java,* Morgan Kaufmann Publishers.

Witten I.H., Frank E. & Hall M.A. (2011) *Data mining : practical machine learning tools and techniques,* Morgan Kaufmann, Burlington, MA.

Figure Legends

Figure 1. Map of lower North Island with aquatic invertebrate sampling sites used to construct and test the Bayesian belief network: blue dots indicate sites in the Manawatu River catchment; red dots represent other rivers and streams in the region.

Figure 2. Screen capture of Bayesian belief network model developed in Netica^{TM}.

Figure 3. Map of river segments in the Manawatu catchment, colour-coded based on the water quality of the predicted QMCI state. Blue = clean, orange = moderate and red = poor water quality, as indicated by the QMCI.

Table 1. Thresholds for states used in BBN nodes. See Appendix 1 for abbreviation definitions and Leathwick *et al.* (2010) for details on how the variables were derived.

	Very High	High	Moderate	Low
USCalc		> 1.64	1.64 > x > 1.48	< 1.48
UsHard		> 2.82		< 2.82
USPhosporus		> 1.77	1.77> x > 1.55	< 1.55
LogNConcen		> 0.1		< 0.1
USPasture	> 0.9	0.90 > x > 0.68	0.68 > x > 0.28	< 0.28
ReachHab		> 4.0	4.0 > x > 3.6	< 3.6
SegRipNat		> 45	45 > x > 20	< 20
SegRipShad		> 0.45	0.45 > x > 0.20	< 0.20
USAvgSlope		> 15.6	15.6 > x > 2.89	< 2.89

Table 2. Metrics of the fit of 300 test sites to five models, for predicting QMCI, created with different algorithms in WEKA from 663 sites. [1] Spherical payoff, a similar but slightly better metric for BBNs than the area under receiver operating characteristic curves (ROC) (Marcot, 2012).

	Percent error	Area under ROC	Cohen's kappa
BBN	28.0	0.760[1]	0.53
Classification tree	38.0	0.729	0.39
Random forests	35.3	0.788	0.42
Artificial neural network (with QMCI classes)	38.0	0.768	0.39
Simple logistic regression	32.7	0.784	0.44

Table 3. Effect of node state on the probability (as a percentage) of clean or poor ecological condition as measured by QMCI based on the developed Bayesian belief network (Fig. 2).

	Ecological condition	
	Clean %	Poor %
Water chemistry (WaterChem)		
Good	69	15
Moderate	57	14
Poor	11	41
Habitat quality (HabitatQual)		
Good	73	14
Moderate	37	11
Poor	20	43
Percentage upstream catchment in pasture (USPasture)		
Low	65	15
Moderate	46	19
High	41	25
Very high	44	25

Table 4. Steps to develop a Bayesian belief network (BBN) to model the effect of the environment on a metric or metrics using software such as Netica™.

Step	Task	Potential methods to achieve task
1	Identify target metric(s)	
2	Identify potential environmental drivers of metric state	Can use statistical (e.g. WEKA as in our study), expert panel, critical thinking, or literature review approaches. Caveat: too many variables will make development challenging.
3	Arrange in influence diagram	Can use software (e.g. Netica™; our study), pencil or whiteboard. For examples see Figure 2 or Allan et al. (2011).
4	Divide all variables into states (i.e. discretise).	Can use statistical approach (e.g. CART; our study), divide into even groups (around 4) or critical thinking.
5	Populate Conditional Probability Tables.	Can use software if suitable data are available (e.g. Netica™; our study), regression equations, or expert opinion.
6	Evaluate BBN model on independent data.	See Marcot (2012).
7	Repeat steps 3 – 6 if necessary to improve model fit to independent data	

Figure 1

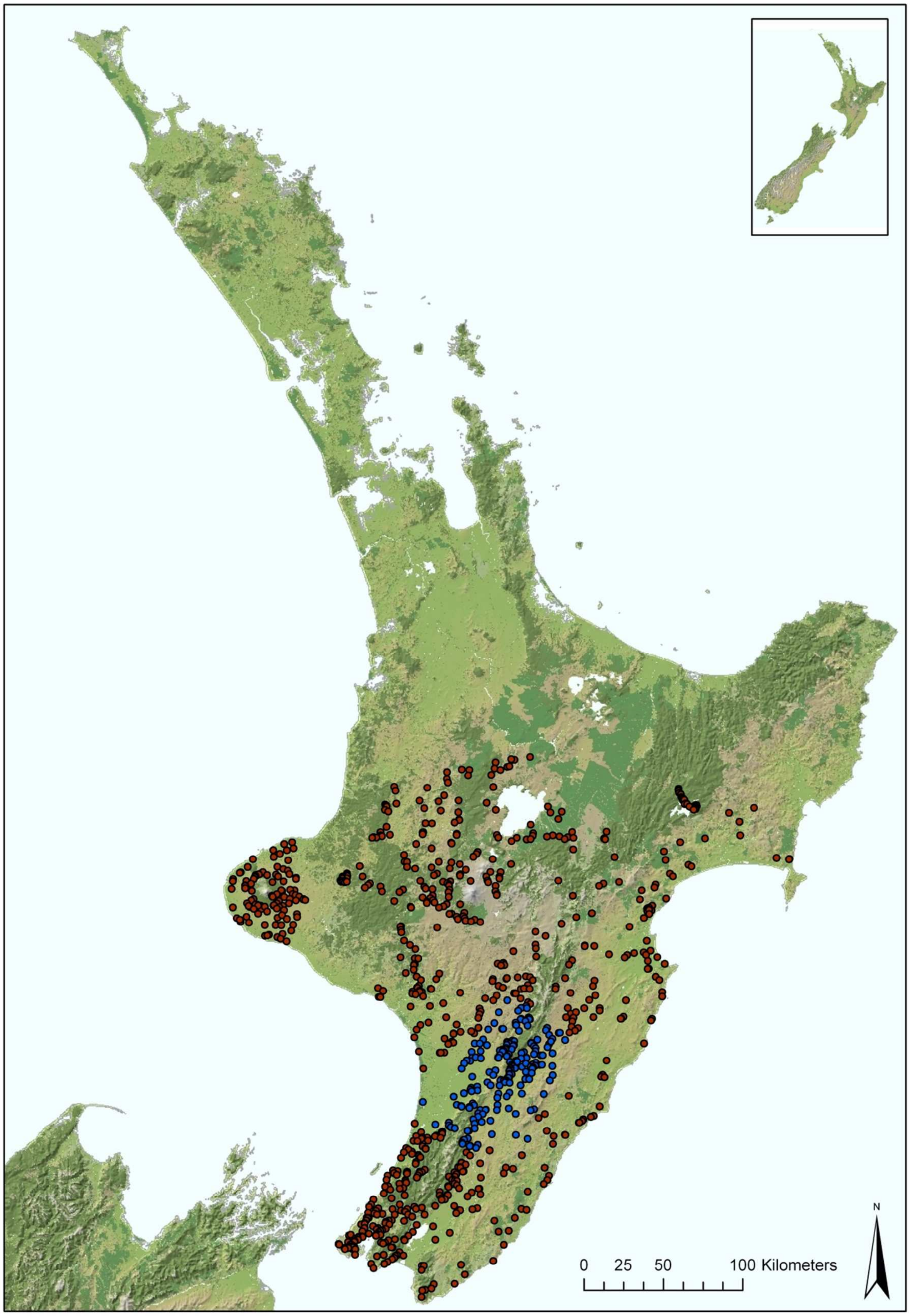


Figure 2

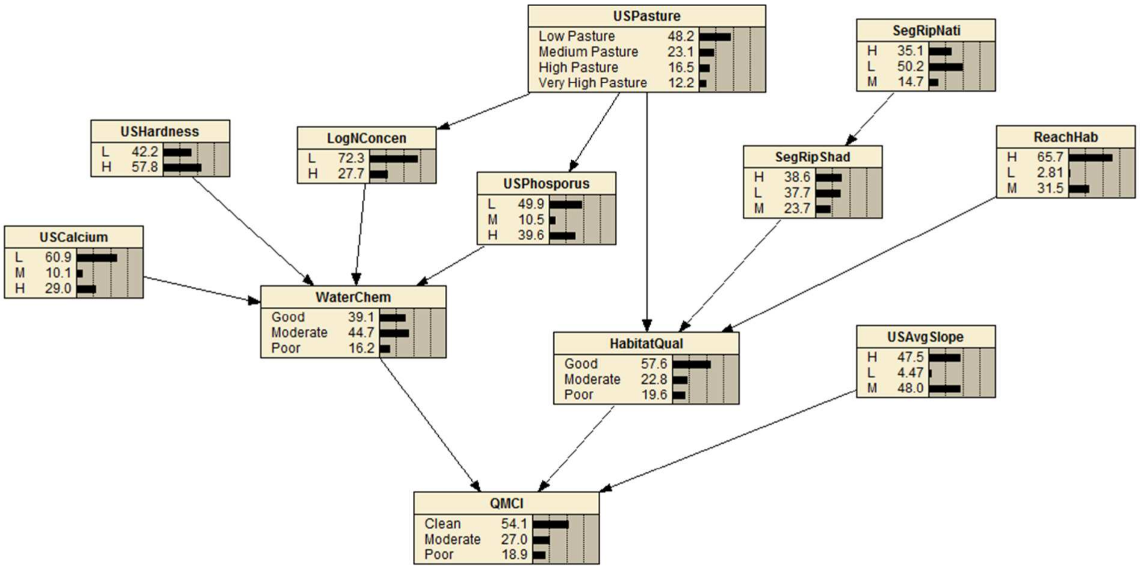
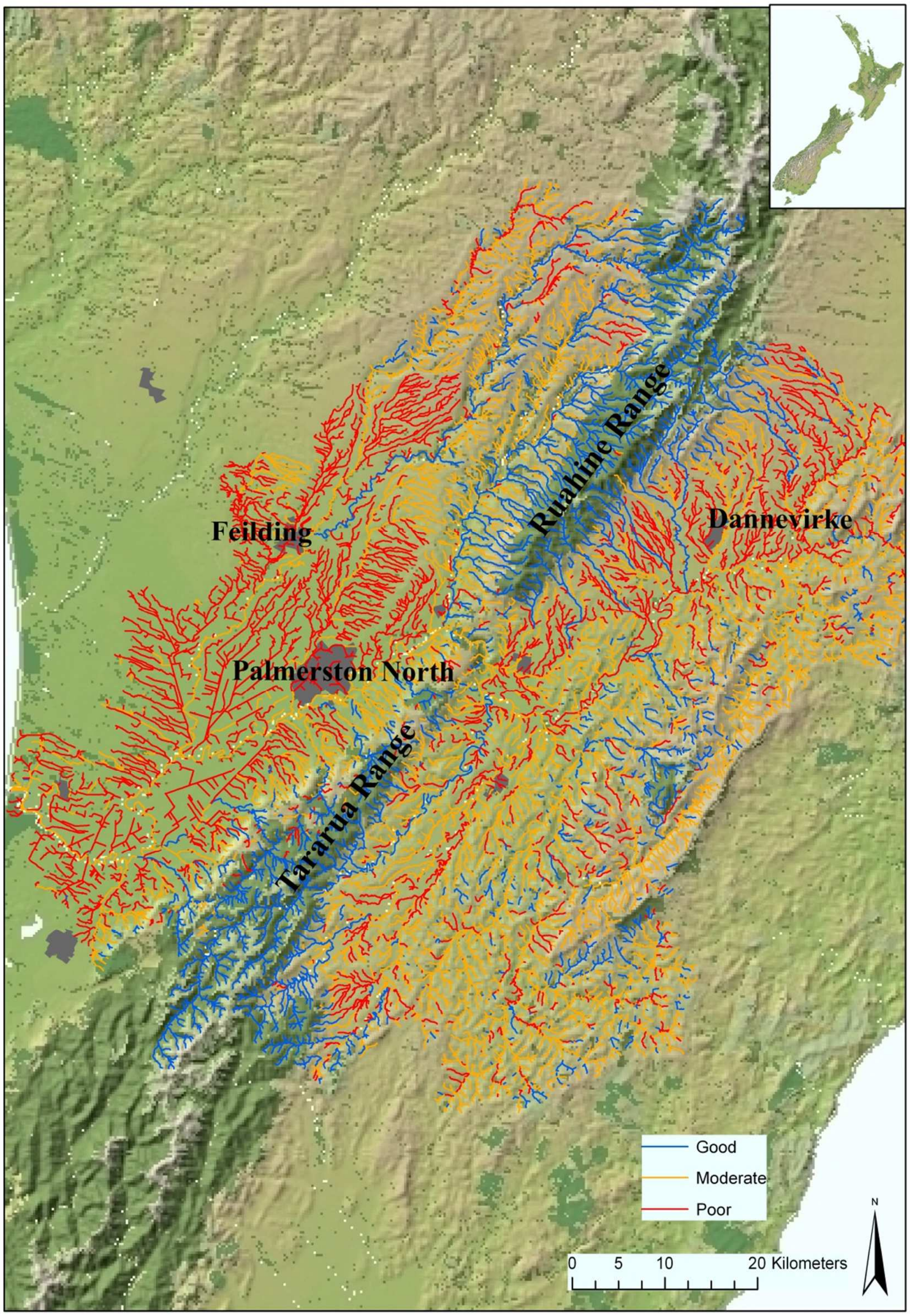

USPasture
Low Pasture 48.2
Medium Pasture 23.1
High Pasture 16.5
Very High Pasture 12.2
SegRipNati
H 35.1
L 50.2
M 14.7
USHardness
L 42.2
H 57.8
LogNConcen
L 72.3
H 27.7
USPhosporus
L 49.9
M 10.5
H 39.6
SegRipShad
H 38.6
L 37.7
M 23.7
ReachHab
H 65.7
L 2.81
M 31.5
USCalcium
L 60.9
M 10.1
H 29.0
WaterChem
Good 39.1
Moderate 44.7
Poor 16.2
HabitatQual
Good 57.6
Moderate 22.8
Poor 19.6
USAvgSlope
H 47.5
L 4.47
M 48.0
QMCI
Clean 54.1
Moderate 27.0
Poor 18.9

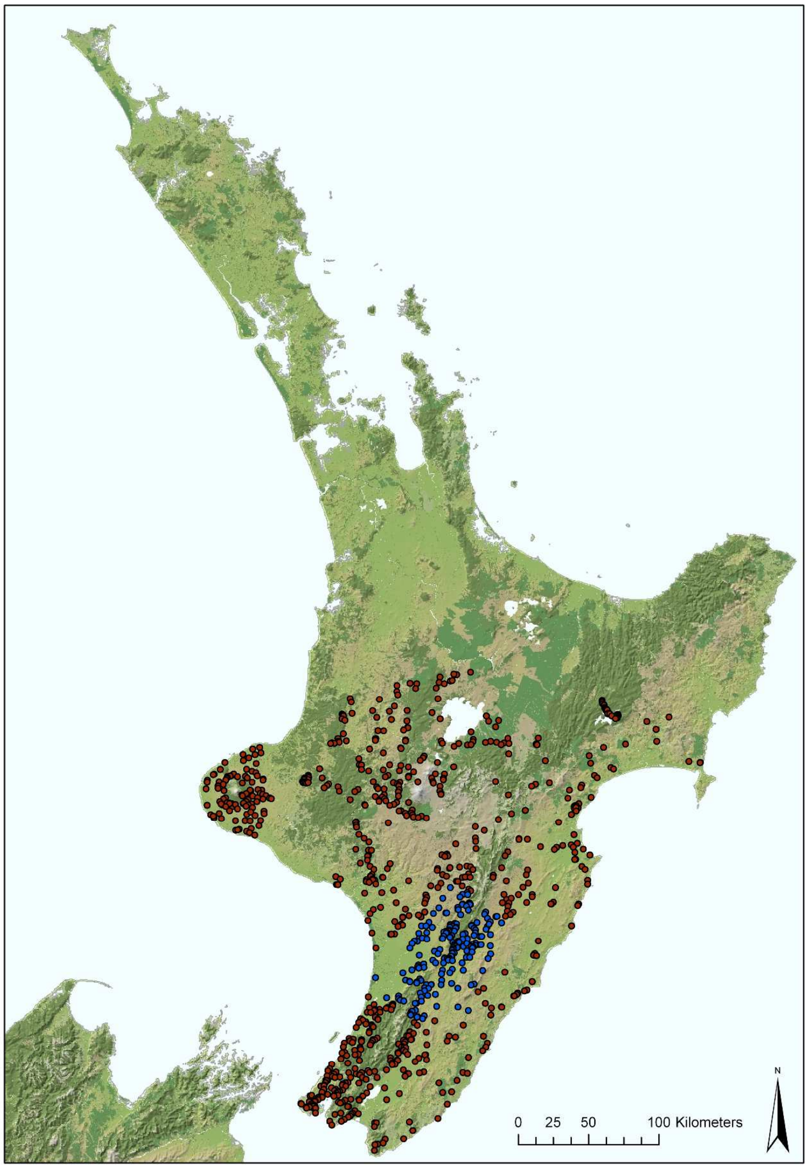
Figure 3


Feilding
Palmerston North
Ruahine Range
Tararua Range
Dannevirke
Good
Moderate
Poor
N
0 5 10 20 Kilometers

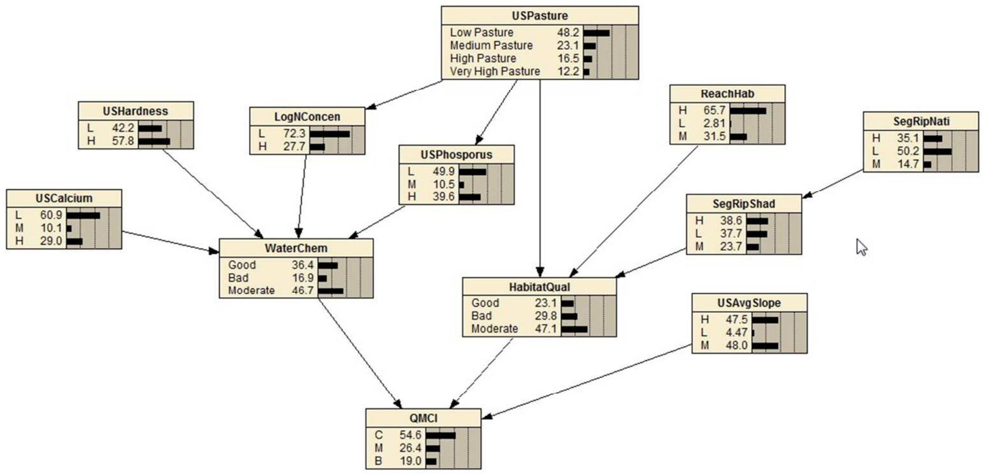
Appendix 1. Variables from the River Environment Classification (REC) (Snelder, 1998; Snelder & Guest, 2000; Snelder & Biggs, 2002; Snelder *et al.*, 2004) or Freshwater Ecosystems of New Zealand (FENZ) (Leathwick *et al.*, 2010) geodatabase used for initial testing of BBN to predict QMCI.

Symbol	**Variable**	**Symbol**	**Variable**
DISTSEA	Distance to sea	DSAVESLOPE	Downstream average slope
CATCHAREA	Catchment area	Q_LRI_2	Flow weighted LRI land class 2
AVEELEV	Average elevation	TOPALLUV	Top rock alluvium
ACCSLOPE	Catchment Accumulated slope	XSINUOSITY	Channel sinuosity
ACCRAIN	Catchment accumulated rainfall	XPASTORALP	Percent pasture
ACCTEMP	Catchment accumulated temperature	Impervious	Percent catchment in impervious surface
ACCEVAP	Catchment accumulated evapotranspiration	NaturalCov	Indigenous vegetation cover in the upstream catchment (proportion)
ACCFLOW	Catchment accumulated flow	LogNConcen	Log10 nitrogen concentration (ppb), range from -4.1 (very low concentrations) to 3.1 (very high concentrations)
CTCHSLOPE	Catchment slope	Downstream	Downstream effects of dams/barriers.
CTCHRAIN	Catchment rainfall	UpstreamDa	Upstream effect of dams/barriers on diadromous species
CTCHTEMP	Catchment temperature	FishEffect	Summed exotic fish effects
CTCHEVAP	Catchment evapotranspiration	Saltru	Predicted probability of capture for *Salmo trutta* (brown trout)
CTCHFLOW	Catchment flow	SumAverage	Pressure indices calculated from individual pressure factors (average)
UPELEV	Upstream elevation	SumMinimum	Pressure indices calculated from individual pressure factors (minimum)
DOWNELEV	Downstream elevation	SegJanAirT	Reach segment January air temperature
RAINALL	Rainfall	SegMinTNor	Reach segment minimum temperature
URBAN	Percent catchment urban	SegFlow	Reach segment flow
FARMING	Percent catchment farming	SegLowFlow	Reach segment low flow
NATIVE	Percent catchment native vegetation	SegFlow4th	Reach segment flow 4th rooted
EXOTIC	Percent catchment exotic forest	SegFlowVar	Reach segment flow variability
SCRUB	Percent catchment scrub	SegSlope	Reach segments slope

TUSSOCK	Percent catchment tussock	SegSlopeSq	Reach segment slope squared
BARE_GROUN	Percent catchment bare ground	SegRipShad	Reach segment riparian shade
TOPEAT	Top rock as peat	SegHisShad	Reach segment historical shade
TOPLOESS	Top rock as loess	SegRipNati	Reach segment riparian native vegetation
TOPALLUV	Top rock as alluvium	SegCluesN	Reach segment CLUES nitrogen
TOPOTHER	Top rock classed as other than classes presented	SegCluesLo	Reach segment CLUES loss
TOPMUD	Top rock as mud	DSDist2Coa	Distance to coast
BASELOESS	Base rock as loess	DSAvgSlope	Downstream average slope
BASEWIND	Base rock as windblown sand	DSAvgSlo_1	Downstream average slope
BASEALLUV	Base rock as alluvium	DSMaxLocal	Downstream maximum local temperature
BASEOTHER	Base rock as other than classes presented	USAvgTNorm	Upstream average temperature
BASEMUD	Base rock as mud	USDaysRain	Upstream days of rain
BASEGREY1	Base rock as greywacke	USAvgSlope	Upstream average slope
LLAKE	Percent catchment as lake	USCalcium	Upstream calcium
UR_Dairy	Upstream catchment as dairy farming	USHardness	Upstream hardness
UR_Beef	Upstream catchment as beef farming	USPhosporu	Upstream phosphorus
UR_Sheep	Upstream catchment as sheep farming	USIndigFor	Upstream indigenous forest
COND	Conductivity	USNative	Upstream native vegetation
PROPSLIP	Proportion of land erosion slips	USPasture	Upstream pasture
Q_LCDB_7	Flow weighted LCDB land class 7	ReachSed	Reach sediment
Q_W_RNVAR	Flow weighted rain variability	ReachHab	Reach habitat quality
Q_W_RD100	Flow weighted rain days greater than 100 mls		



Map of lower North Island with aquatic invertebrate sampling sites used to construct and test the Bayesian belief network: blue dots indicate sites in the Manawatu River catchment; red dots represent other rivers and streams in the region.
296x419mm (300 x 300 DPI)



Screen capture of Bayesian belief network model developed in NeticaTM.
383x194mm (72 x 72 DPI)