

Exploratory use of a Bayesian network process for translating stakeholder perceptions of water quality problems in a catchment in South Africa

NA Rivers-Moore^{1*}

¹Centre for Water Resources Research, University of KwaZulu-Natal, P/Bag X01, Scottsville, 3209, South Africa

ABSTRACT

Water resource management is complex, and should ideally be a co-operative, stakeholder-driven problem-solving process. Bayesian networks (BNs) are one participatory tool being increasingly used to facilitate this process. The upper Mgeni catchment in the province of KwaZulu-Natal, South Africa, is a key water resource area with looming water quality problems. The high number of stakeholders involved in a catchment management forum provided an environment for testing the development of a BN showing relationships between water quality problems and stakeholders in this area. Through engagement with stakeholders at quarterly meetings during 2011, and collation and analysis of water quality time-series, qualitative and quantitative data were successfully translated into a BN for water quality improvement in the study area. The model demonstrated that certain water quality variables (for example, compliance of wastewater treatment works; increase in housing developments) were more likely to be the cause of problems than others (such as discharges from farm dams or accidental spills). The value of involving stakeholders in a co-operative process is highlighted, and it is argued that the robustness of such a model would be enhanced further if applied within a formal participatory approach using conditional probability values endorsed by all stakeholders.

Keywords: catchment management, land use, probability, stakeholder engagement, water quality

INTRODUCTION

Water resource management is complex, involving multiple stakeholders and numerous interacting variables. Typically, maximum allocation of a limited resource in a multi-stakeholder environment results in competitive behaviour to secure sector-specific portions of this resource, which is a less-than-equitable outcome for downstream users. These are often reactive, crisis-driven approaches, where what is ideally required is a co-operative, stakeholder-driven problem-solving process. Even assuming problem solving occurs in a co-operative environment, solving such puzzles in the absence of suitable conceptual tools is sub-optimal because typically human short-term memory can only consider up to 7 points (Bacon et al., 2002). Additionally, management plans often fail as they are developed without participation of local individuals (Cain et al., 2000). It is becoming increasingly obvious that stakeholder participation is critically important in successful water resource management (Zorilla et al., 2010). There is growing recognition of the contribution of 'citizen science' to promote public engagement in environmental issues (Dickinson et al., 2012), but also of the large role of tacit knowledge in the decision-making process. Incorporating such less-accessible knowledge into the decision making process has value, even though it is recognised that decisions are typically based on perceptions rather than absolute reality (Zorilla et al., 2010). This is unfortunate because perceptions may be based on limited experiences that falsely inform intuitive thinking, and downplay more objective data (base rates or prior probabilities – see Kahneman, 2011).

Bayesian networks (BNs) are knowledge representations that can be used in situations involving reasoning and

decision-making under uncertainty (Kjaerulff and Madsen, 2008). Uncertainty implies that the quality or state of a variable or decision outcome is unknown: where multiple stakeholders are involved, co-reasoning is required to generate options and consider the consequences of each option based on limited data, which should precede decision making. BNs provide a method of representing relationships between variables, even if the relationships involve uncertainty, unpredictability or imprecision (Batchelor and Cain, 1999). They are a useful tool for visualising complexity and engaging stakeholders (Zorilla et al., 2010). BNs have been used successfully as a participatory tool in representing complex issues in a number of environmental studies, and as a decision support tool to evaluate different management options (Ames et al., 2005; Lynam et al., 2010; Shenton et al., 2010; Stewart-Koster et al., 2010; Webb et al., 2010). Increasingly, alternatives to deterministic models to inform decisions are enabling a more creative stakeholder-driven discursive approach to solving complex problems in connected systems (Cain et al., 2000; Zorilla et al., 2010; Eröset al., 2011). Key in choosing such tools are their capacity to capture perceptions, and to utilise non-scientific information in a structured, logical form, that also considers base rate data (prior probabilities or base class probabilities).

The complexities of resolving water quality issues in South Africa's rivers, in spite of relatively good data, continues to hamper problem resolution, and in spite of a growing urgency to do so. The Mgeni River, a key water resource for the province of KwaZulu-Natal, eastern South Africa, has been identified as a system whose water quality poses potentially serious health risks to people and agriculture (CSIR, 2010). In a recent report-back to South Africa's Parliament on water quality, the Mgeni River was singled out as a system with growing water quality problems due to 'poultry farms, effluent from cattle feedlots and informal settlements without sanitation facilities' (Anon., 2011).

* To whom all correspondence should be addressed.

e-mail: blackfly1@vodamail.co.za

Received: 26 February 2014; accepted in revised form 16 March 2016

The upper Mgeni catchment has been subject to human activity since at least the 15th century, when Black pastoralists lived in the area. From 1850, White farmers settled the area (Moll, 1965). Since the most extensive natural climax community in the catchment is *Themeda triandra* grassland, noted for its good grazing (Moll, 1965), livestock agriculture became the dominant land use, and by 1970 dairy dominated as the key agricultural activity (Scotney, 1970). While wetlands were historically a notable feature, there has been a slow attrition of wetlands over time, due to increased land use pressures, and, since 1944, the catchment has undergone increasing fragmentation, urbanisation and impoundment (Rivers-Moore, 1997). Accelerated trends in land-use change, primarily from agricultural to residential (Kallicharan, 2010), are likely to have water quality impacts. A gradual attrition of agriculture and dairy farms to other types of land use has culminated in a situation of cumulative housing developments (low-income and luxury) whose wastewater contributions far exceed the existing capacity of the current infrastructure. These issues have been recognised but not entirely quantified or integrated by concerned catchment stakeholder discussion groups.

The aims of this research were twofold: firstly to integrate stakeholder discussion points into a quantifiable framework through the development of a BN for key variables affecting water quality in the upper Mgeni catchment and, secondly, to demonstrate the use of a BN as a tool for facilitating negotiations between various catchment stakeholders. The focus of this study was centred on concerns with point-source pollution affecting water quality, particularly as this links to human drinking water quality as a function of municipal management efficiency and defined from the perspective of catchment

stakeholders. Consequently, non-point source pollution was not considered in this study.

METHODS

Study area

The study area fell within one of the upper quaternary catchments (sensu Midgely et al., 1995 – management units defined on a standardised runoff per unit area) of the Mgeni River (Fig. 1). The rolling nature of the landscape limits intensive agriculture (Moll, 1965; Scotney, 1970), but favours extensive agriculture. Midmar Dam, fed by a number of tributaries but primarily by the Mgeni River, dominates the area. Grassland and natural vegetation (non-grassland Mistbelt mixed *Podocarpus* forest, bushland and woodland) make up over 50% of the catchment, while land use generally consists of a mix of forestry and agriculture (mixed crop cultivation and stock farming – poultry, dairy and piggeries). The abundance of dolerite in the catchment, and associated dykes, together with a high rainfall (Scotney, 1970), explains the large number of wetlands in the catchment. Also present in the study area are rural and urban settlements (current and proposed).

Water quality issues and stakeholder groups

Engagement with stakeholders and identification of key water quality issues began with, and focused considerably on, building a relationship with members of the Upper Mgeni Catchment Management Forum (CMF), for whom this area is a particular focus. CMFs are facilitated by the Department

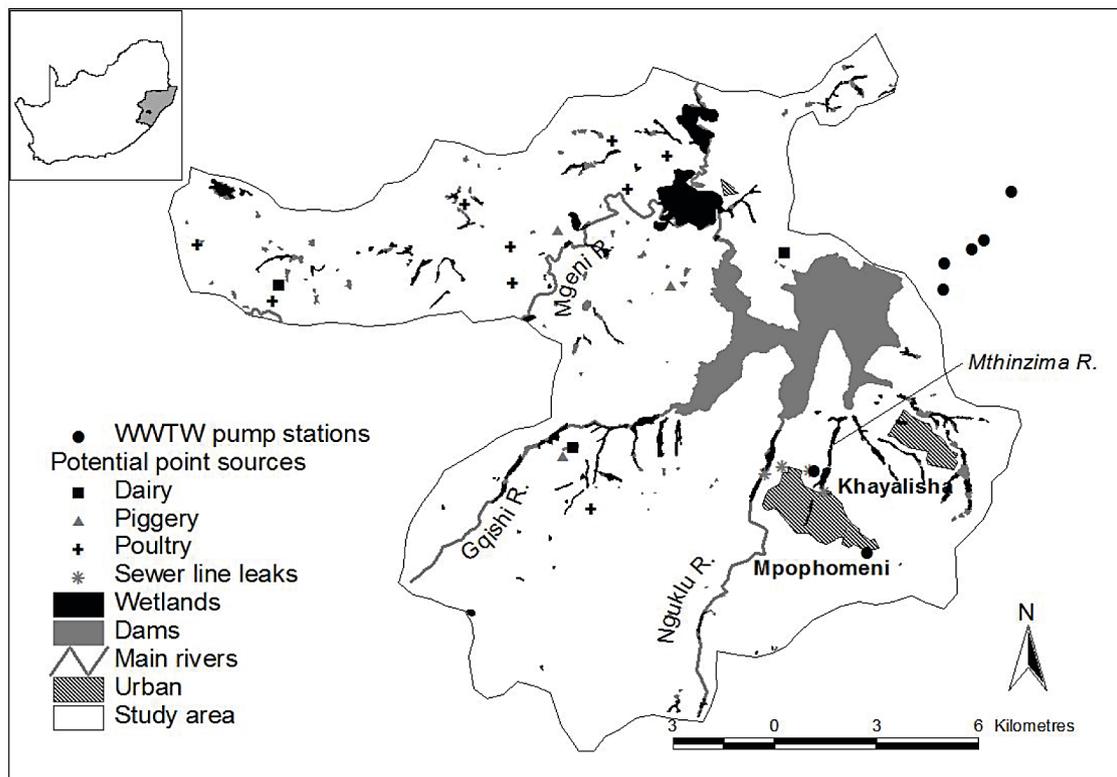


Figure 1

Quaternary catchment U20C, showing salient features used in the development of the BN to facilitate water quality decision making. The Mthinzima River (not shown) is to the east of the Nguklu River, flowing out of Mpophomeni and through a degraded wetland. All three rivers flow into Midmar Dam, with the town of Howick, as the main economic centre for this area, approx. 2 km to the east (not shown).

of Water and Sanitation, through provision by the 1998 South African National Water Act. These forums are in place to facilitate information and feedback to community and stakeholders, and provide a platform to discuss catchment-related issues. Members of the upper Mgeni CMF include, inter alia, individuals from conservation non-governmental organisations (NGOs, including the Dusi-Mgeni Conservation Trust: DUCT), community members from urban and rural areas, Umgeni Water (regional water supplier), residents' associations, and farmers. Attendance is voluntary, and meetings are typically attended by 15–20 people who all have some degree of interest in the local catchment as either residents or stakeholders.

Three meetings were attended (24 February, 26 May and 24 November 2011), where presentations and interactive exercises were undertaken. The concept of BNs was new to CMF members, who represent a range of socio-economic and educational backgrounds. Other techniques of qualitative research were not used for this study, as its focus was to assess whether discussions emerging from quarterly meetings could be quantified and integrated using a BN approach. In the first meeting, the process was initiated by demonstrating the ease with which stakeholder conflict can occur between competing catchment users. This was achieved using a catchment role-playing exercise, based on an example from Cain (2001). Here, CMF members were divided into 4 stakeholder groups representing commercial agriculture, subsistence agriculture, urban area and conservation. Each group was provided with a mandate to maximise resource use for their sector.

Next, BNs were presented as a parallel approach for representing the same issues, and as a tool for co-operative rather than competitive problem solving. CMF members were questioned on whether they saw this approach as useful. After this process, informal discussions were held to understand what the key water quality issues were. These issues were verified and expanded on after the first meeting, through a process of reviewing previous CMF minutes, popular articles in local newspapers, and reviewing extensive transcripts of e-mails between DUCT members during 2012. Further insights on more complex issues were gained from interviews with regional water quality specialists (Terry, 2012), key members of DUCT who are involved with policing municipal water quality compliance as 'citizen scientists', and an agricultural sector stakeholder (Howick District Land Owners Association). Through all of these exercises, a large amount of qualitative information was gathered.

Between the first and second CMF meetings attended, an initial BN was developed by the author, which was presented at the CMF meeting of 26 May 2011 for comment from the CMF members. This was then further refined, and populated with probabilities, before being presented at the third CMF meeting attended, on 24 November 2011.

Construction of the Bayesian network

Bayes' theorem provides a mathematical framework for the updating of belief with new evidence (Jensen and Nielsen, 2007; Stewart-Koster et al., 2010). BNs do not necessarily refer to causality, and do not require that the links represent causal impact, but rather enable one to follow how a change of certainty in one variable may change the certainty for other variables (Jensen and Nielsen, 2007). However, in this study application, relationships were constructed to be causal. A BN is thus a type of graphical model where variables have mutually exclusive states, and information is transmitted between parent and child nodes using Bayes' theorem and the chain rule (Jensen and Nielsen,

2007). Updating of prior probabilities based on new information does not necessarily occur. A BN is a directed acyclic graph (DAG) with elements (V, E) (where V = variables and E = events or states), and representing relationships between variables. There are two steps to defining a BN (Kjaerulff and Madsen, 2008; Jensen and Nielsen, 2007):

- Qualitative component: identification of variables, states (events) and relationships between them = DAG + states
- Quantitative component: knowledge on (usually) causal relations, conditional (joint) probabilities and utilities (preferences) associated with decision options

Each node has system states, and the state of the child node is conditional upon the states of its parent/s nodes, with that relationship defined by conditional probability tables that may be derived using either qualitative or quantitative data. Nodes in a BN are of 3 possible types (chance, decision and utility), which can either be discrete or continuous. Once constructed, the parameters of a BN may be continuously updated with new information (Fig. 2). The addition of decision variables and utility nodes (costs/benefits) further facilitates the decision-making process as the decision with the highest expected utility is indicated (Kjaerulff and Madsen, 2008).

An initial step in this application was to define the critical season to which it would be best applied. This was considered to be when runoff from the catchment was highest, because high runoff was likely to coincide with high flows having to be processed by the wastewater treatment works (WWTW), which is typically when these systems fail. An alternative season, not chosen for this study, would be the low-flow season, as low-flow situations are often more sensitive to cumulative water quality stressors (Nilsson and Renöfält, 2008; Mantel et al., 2010) because dilution capacity is smallest. However, this was seen to apply more to catchment-wide water quality issues rather than to impacts of municipal management-related point-source water quality issues, as per the objectives of this study. Forty years of mean daily flow data (2 December 1948 to 1 December 1988) were split into two unequal parts (1949–1962 and 1963–1989) to compare mean daily flow rates for each month before and after impoundment due to completion of Midmar Dam in 1963. Flow data for the nearest gauging weir (U2H001) with extensive flow data were used for analyses. The baseflow

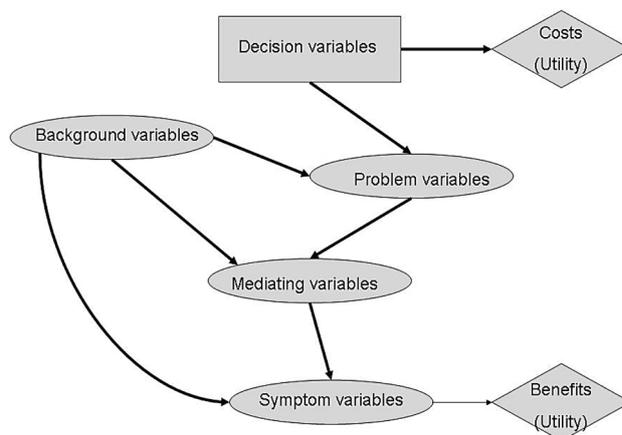


Figure 2

Diagram of key components of a Bayesian network (after Kjaerulff and Madsen, 2008: 153)

index of the flow data was calculated using suitable software (Nathan et al., 2007) to provide an indication of the possible relative inputs of surface flows versus groundwater inputs to assess final water quality.

Next, the qualitative components of the BN were developed by defining variables and states using an iterative process. Based on the current understanding of the issues in the catchment, variables were defined to represent cause-and-effect issues. Variables were categorised into chance (background, problem, mediating and symptom – Fig. 2), decision and utility categories, and states were assigned to each variable. All variables used discrete states, arranged from most positive to most negative, with two states assigned to each variable, apart from one variable which had three. Nodes were linked in cause-and-effect sequences using the Bayesian software Netica v 4.16 (Norsys, 2010). Development of a BN was an iterative process of testing the logic of relationships and keeping the network as parsimonious as possible. In this study, no more than 4 parent nodes were linked to any child node, because the elements of a conditional probability table increase exponentially according to i^n based on number of states (i) and the number of parent nodes (n) (Cain, 2001).

Data sources and population of diagram

The 'quantitative' aspect of building the BNs, i.e., defining probabilities and populating conditional probability tables, was achieved using a combination of expert knowledge and opinions (qualitative input data) and probabilities calculated from time-series (quantitative input data). For the former approach, probabilities were elicited after discussions with key stakeholders, as described in the previous section, and numbers assigned based on numbers of belief (as per Kjaerulff and Madsen, 2008), i.e., qualitative judgements (unlikely versus probable) are translated into probability values (0.1 versus 0.8 probabilities).

Base probabilities for certain parent variables (for example, cable theft) were calculated from DUCT reports on water quality from January to November 2011. These were circulated by DUCT as a spreadsheet with the date of a water quality related issue, and the description of the issue. From this information, events were categorised (spill, manhole leak), with associated return intervals (and hence probabilities) calculated using the number of events per category, and the number of sites at which an event occurred.

Time-series data from which return intervals could be calculated based on threshold values were obtained from either published data (rainfall – Lynch, 2004) or databases maintained by the regional water service supplier based on ongoing monitoring programmes (Terry, 2012). Very few published datasets on water quality in the upper Mgeni catchment were available, with the only formalised report to the knowledge of the author being a 'State-of-Rivers' report (WRC, 2002), that identifies nutrient input from dairies and piggeries, and faecal contamination from informal settlements, as particular issues. However, these are not quantified, and it was considered most appropriate to obtain information on water quality issues from expert opinion. Rainfall was included as a water quality risk due to the link between high-intensity rainfall events and runoff. Daily rainfall data were available for the Mpophomeni area (SAWB station no. 239102) for an 8-year period from 1 August 1991 to 15 October 1999, for which a return interval for 'high' rainfall events (defined in this study as > 3 times the daily rainfall average on days when rain

fell over the overall record length, on the basis that this is a relatively large-magnitude rainfall event but with a reasonable return interval period of 60 days) was calculated. Water quality data (temporal resolution ranged from weekly to monthly to less frequent, and generally for the period 1990–1999/2011) was extracted for 6 monitoring sites which were selected based on the catchment land uses they represented. Since no single metric on its own provides a reliable estimate of whether water quality is good or bad (Terry, 2012), 3 metrics were used, viz., *E. coli* count, conductivity and nitrates. For this study, nitrates were chosen in preference to phosphates because the former are more mobile within a catchment, and because nitrate levels reflect effluent inputs from oxidised ammonia. Nitrates enter both surface and groundwater as a result of, inter alia, agricultural activities and wastewater disposal. Return intervals of exceedances of thresholds for drinking water for each metric were calculated: *E. coli* = 1 000-100 mL⁻¹, conductivity = 30 mSm⁻¹ and nitrates (NO₃) = 6 mg·L⁻¹ N (all as extrapolations of drinking water guidelines from DWAF 2005 and WHO 2011, and adapted to include distinction in return intervals between sites for the water quality data used in this study). Here, it is acknowledged that a threshold for *E. coli* of 1 000-100 mL⁻¹ is in excess of recommended drinking water quality standards; for the purposes of this study any threshold below this number would have yielded similar return intervals at all sites and made comparisons meaningless. Final probabilities of whether water quality was 'good' or 'bad' per land use were calculated as the sum of all three individual probabilities of threshold exceedance, on the assumptions that metrics are independent, and that the individual metric probabilities are less robust than their sum as a measure of poor water quality.

Conditional probabilities of child nodes were calculated using software developed by Cain (2001). It is acknowledged that techniques are available to populate conditional probability tables (CPTs) that include an elicited process with stakeholders, population by experts, or in-filling based on observations collated into a case file used by learning algorithms in Bayesian software. These techniques were not used in this study because the aim of this research was to ascertain whether wide-ranging discussions within a catchment management forum meeting could be integrated into a simple model that illustrated the inter-connectedness of seemingly disparate issues. The approach of populating CPTs using reasonable numbers is an acknowledged weakness of this study, even though the defensibility of these numbers was not seen as a primary goal of this study.

A decision node and two utility nodes were added to assist with the decision-making process of evaluating the impacts of different variable states. Decision nodes reflect states of possible interventions, while utility nodes reflect the 'value' (cost versus benefit), either in monetary or relative terms, of different outcomes (Jensen and Nielsen, 2007). A single decision node was added to include management actions to capture the value of environmental resilience, and management accountability. The relative value of each option was assessed based on assigned relative utility (benefits) values of 0 and 1 for 'poor' and 'good' water quality respectively. A cost utility node associated with the decision node was added, where the 'do nothing', 'management accountability' and 'rehabilitate wetland' actions had relative cost scores of 0, -0.1 and -0.2 respectively. Scenarios were assessed based on combinations of states for selected variables and the resultant management scores based on the utility values.

RESULTS

Identified issues

Analyses indicated that the period of highest flows has been delayed by one month, from February until March, as a result of managed flow releases from an upstream impoundment. In this study, the critical season where the outputs from the BN would be of most relevance was consequently identified as February–March. The annual baseflow index (ratio of mean annual baseflow to mean annual flow) of U2H001 is 0.783, which indicates a stable discharge with relatively high dependence on groundwater for the Mgeni River.

Based on recurring concerns raised at the catchment management forum, and discussions with various stakeholders, the following issues emerged that formed that basis for developing the BN:

Agriculture

Members of the CMF voiced concerns over nitrates and phosphates from agriculture going into rivers. A number of farms with intensive dairy, piggery and broilers (poultry) occur in the catchment. Excrement waste from these land uses is typically transferred to settling ponds before being applied as irrigated fertiliser to crops. Leaking effluent dams represent a loss of income to a farmer, as fertiliser then needs to be bought, but effluent dams still need to be well managed to prevent loss to streams in rainfall-runoff events. Impacts of forestry on water quality have not been considered in this study.

Current urban areas

The urban/ peri-urban area of Mpophomeni was built on a dairy farm that was expropriated by the Government in 1963. This area continues to expand, with more formal houses towards the lower part of the catchment, and poorer, less formal houses located in an economic gradient extending up the catchment.

Cumulative developments

A large number of individual developments are planned or currently being built in the catchment. The general consensus of stakeholders is not so much concern about the impact of each development, but rather the collective impact of all developments taken cumulatively. Concerns relate to the capacity of sewage plants to handle effluent with an increase in housing developments.

Wastewater treatment works

Sewage is pumped uphill from the defunct Mpophomeni WWTW to the Howick sewage treatment plant, approximately 8 km northeast. Reticulation infrastructure includes 7 pump-stations. Probabilities for the WWTW not being compliant according to South Africa's Department of Water Affairs Green Drop certification process for effluent standards were not available for the Mpophomeni WWTW, and the closest functional WWTW was used (Howick: reported compliance of 83% for 2011; DWA, 2011: 189). Failure in the system may result in water quality problems: power cuts, vandalism and theft of power cables leads to failure of pump stations and overflows of sewage; inconsistent cleaning of screens which intercept non-liquid matter causes

pump-station problems; blockages and leaks in the sewer reticulation system; and overflowing manholes that discharge sewage into streams. Households may rely on pit latrines and septic tanks that cumulatively impact on groundwater quality.

Miscellaneous environmental issues

Concerns were raised over groundwater contamination due to inappropriately placed cemeteries, and leaching from landfill sites (formal and informal). Stormwater pipes which are fed by runoff can also leach into groundwater, or enter sewer reticulation systems. Cumulative developments exacerbate these problems. Random events of spills (for example, an accidental chemical spill into a river at a particular point) add to surface water pollution.

Community involvement

Not quantifiable but discussed at the CMF has been the role of 'sense of place' in the stewardship of an area. This manifests through local residents participating in community gardens and wetland clean-ups, church groups, and a sense of community. In the absence of municipal management self-regulation and accountability, the chances of outright management failure are reduced through the influence of pressure groups, such as NGOs (DUCT, CMFs). Public awareness and alerting relevant individuals to water quality issues all play some role in reducing the likelihood of reduced water quality.

Management options

Three management options were identified in this study. The 'do nothing' option assumes that the status quo continues, and that the trend towards urban development and away from an agricultural economy continues. The 'management accountability' option is difficult to quantify, but it is assumed that consequences for good or bad management decisions are reflected in the quality of water. Here, management accountability was equated with maintaining an agricultural economy linked to promotion of sound land-use practices. Improved management can be facilitated through forums and better communication, audits, prosecution, vision, and co-operation to avoid duplication of effort. Thus, building community awareness is enhanced through involvement of NGOs such as DUCT, and creates a channel to put pressure on municipal managers and to alert people to water quality problems. Valuing environmental services in improving water quality is achieved through wetland rehabilitation, which builds environmental resilience. Here, it is acknowledged that trapping of pollutants is proportional both to the overall area of individual wetlands, and the cumulative impacts of all wetlands within a catchment (Ellery et al., 2009).

Based on an understanding of the key catchment issues, and desirable versus non-desirable conditions, the qualitative part of developing a BN was achieved by defining 17 variables with generally 2 states each (Table 1).

Calculation of probabilities and derivation of conditional probability tables

Prior probabilities for parent nodes were estimated from a combination of expert opinion, return intervals of events based on reported incidents by DUCT, and use of elicited probabilities (Table 2). The probability of a leak in the sewer reticulation system was difficult to assess. Based on the reports of manhole

Variable	States	Category	Comments
Cable theft	No, Yes	Background	WWTW pumps cannot function when electricity cables are stolen
Cemeteries	Good, Poor	Problem	Badly located cemeteries can result in water quality problems due to leaching into groundwater and/or runoff into streams
Dairy/ Pig farms	Sealed, Leaks	Background	Relates to effluent storage dams from dairy or pig feedlots
Environmental resilience	High, Low	Mediating	The ability of the environment to clean itself, i.e., improved water quality as a result of environmental services provided by ecosystems such as wetlands.
Groundwater	Good, Poor	Problem	
Housing developments	Current, Up 10%, Up 20%	Problem	
Management effectiveness	Effective, Not effective	Mediating	
Management options	See text	Decision	
Rainfall	Low, High	Background	A higher-than-average rainfall event will result in increased inflows to WWTW and into the stormwater drainage system.
Screens	Cleaned, Not cleaned	Background	
Sewer reticulation	Contained, Overflow	Problem	Includes full system of sewerage pipes and overflows of manholes.
Spill	No, Yes	Background	A random event resulting from a spill of contaminants ending up in a river
Stormwater	Contained, Leaches	Problem	Stormwater drains and water, landfill and septic tanks lumped into one category.
Surface water	Good, Poor	Problem	
Utility (Benefits)	N/A	Utility	
Water quality	Good, Poor	Symptom	
WWTW	Compliant, Non-compliant	Problem	

leakage events, there were 10 events reported at 9 sites over a period of 11 months, which gave a 91% probability of a manhole leaking somewhere in the catchment each month. For the Mpophomeni catchment, it was assumed that there are a minimum of 50 manhole covers, such that the probability of a particular manhole leaking was 0.91/50 or 2%.

Calculation of return intervals was a more quantitative approach to determining probabilities. Since high-rainfall events are more likely to lead to effluent dam leaks and overflows at the WWTW, it was necessary to define and then estimate the probability of a high-rainfall event. The return interval for rainfall was set at roughly 3 times the mean daily rainfall of 7.82 mm (for days when a rainfall event occurred for the period assessed), or approx. 24 mm. This resulted in a return period of 58.7 days, so that any 'high' rainfall event had a probability of 1.7%.

In the initial engagement process with the upper Mgeni CMF in February 2011, the stakeholders set the objective of wanting 'good' water quality. This needed to be defined in a measurable way in order to set probabilities for this. Additionally, it was desirable in this study to be able to measure the improvement in water quality in response to identified management interventions. Since these data were not directly available, it was necessary to measure this indirectly, which was achieved using water quality time-series data from 5 water quality monitoring points, located in catchments covering a range of land uses. Return intervals for exceedances of reasonable thresholds for *E. coli* (Fig. 3), conductivity (figure not shown) and nitrates (figure not shown) were calculated, with the final probability of 'bad' water quality as the sum of the probability of exceedance for all three

Variable	States	Probabilities (%)
Housing developments	Current, Up 10%, Up 20%	(70, 20, 10) vs. (5, 40, 55)#
Location of cemeteries	Good, Poor	(90, 10) vs. (50, 50)#
Spill	No, Yes	(55, 45)
Dairy/ pig farms	Dam sealed, Dam leaks	(99.5, 0.5)*
Screens	Cleaned, Not cleaned	(90, 10) vs. (55, 45)#
Stormwater	Contained, Leaches	(44, 56)
Sewer reticulation	Contained, Overflows	(98, 2)
Cable theft	No, Yes	(91, 9)

metrics (as derived from return intervals), on the basis that no single metric on its own provides a reliable estimate of whether water quality is good or bad. At all sites, return interval curves showed a successive increase in exceedances as land-use intensity increased. Thus, the 'natural' site had a 5% chance of readings showing poor water quality independent of flow periods. The site downstream of agriculture had a 19% chance of indicating poor water quality, while the site just downstream of Mpophomeni

had an 88% probability of any reading reflecting poor water quality. The probability of poor water quality dropped by almost 15% in the same stream downstream of a degraded wetland, where degradation refers to altered vegetation communities, canalisation and resultant altered wetland hydrology (Table 3). The rural site was in the upper Mpophomeni catchment, and potentially reflects the probability of poor water quality as a result of pit latrines. The odds of poor water quality were calculated against the baseline 5% of the natural site, such that, for example, a rural site with septic tanks was 5 times more likely to have poor water quality than a 'natural' site. In each instance, water quality results are most likely to be a combination of upstream and local conditions, ameliorated to some extent by the distance of the sampling station from upstream impacts.

Probabilities for the environmental resilience node have been derived as follows: 'management accountability' has been equated with 'maintaining an agricultural economy' (high vs. low resilience = 80.80 vs. 19.20%); 'rehabilitate wetland' has been equated with the 'probability of poor water quality for urban + wetland' (high vs. low resilience = 27.65 vs. 72.35%); and 'do nothing' with 'status quo of urban land use water quality' (high vs. low resilience = 11.85 vs. 88.15%).

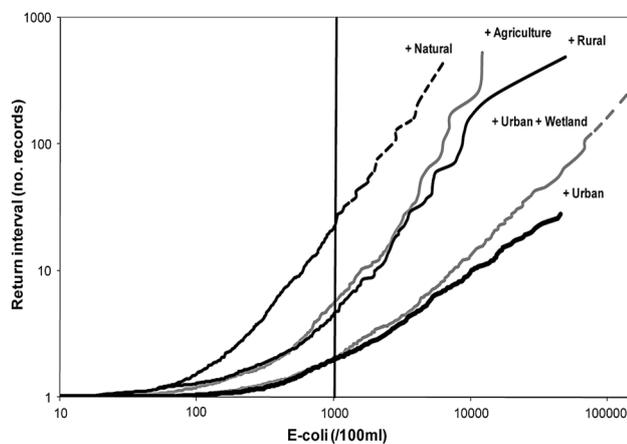


Figure 3

Return intervals for E. coli at 6 water quality monitoring points in the upper Mgeni catchment, covering 5 different land uses

Bayesian network

Based on concerns of stakeholders at the first upper Mgeni CMF attended in February 2011, a preliminary BN consisting of 9 variables was constructed and populated with probabilities. This was refined through successive iterations with the stakeholders, and the final BN had 17 nodes (Fig. 4), with associated probabilities. A simple table of conditional probabilities for the sewer reticulation node, which has parent nodes 'cumulative housing developments' and 'stormwater', is included for illustrative purposes (Table 4).

A management node with 3 management options, and a utility node linked to water quality, with utility values assigned for 'good' and 'poor' water quality (Fig. 4) were added to the BN to assist with the decision-making process. Mediating nodes were added to simplify the conditional probability tables for WWTW, groundwater and surface water. The first mediating node, 'management effectiveness', had probabilities of effectiveness which improve as a result of management accountability but decrease for a 'do nothing' option. In the second mediating node, the impact of rehabilitating the wetland, however, was incorporated into the 'environmental resilience' node, where the action of rehabilitating the wetland increased the probability of high environmental resilience. In this example, with no node states actively selected, the likelihood is that water quality will be 'good', and the most appropriate management intervention, based on utility scores, is to rehabilitate the wetland.

TABLE 4
Conditional probabilities for sewer reticulation being contained or overflowing, based on parent nodes of 'cumulative housing developments' (3 states) and 'stormwater being contained or leaching'

Parent nodes		Sewer reticulation	
Housing	Stormwater	Contained	Overflow
Current	Contained	98	2
Current	Leaches	70	30
10pc	Contained	60	40
10pc	Leaches	43	57
20pc	Contained	40	60
20pc	Leaches	2	98

TABLE 3

Calculation of probabilities of bad water quality (i.e. return intervals of exceedance of thresholds) at 5 monitoring sites under different land uses. Probabilities for the environmental resilience node have been derived as follows: 'management accountability' has been equated with 'maintaining an agricultural economy' (high vs. low resilience = 80.80 vs. 19.20%); 'rehabilitate wetland' has been equated with the 'probability of poor water quality for urban + wetland' (high vs. low resilience = 27.65 vs. 72.35%); and 'do nothing' with 'status quo of urban land use water quality' (high vs. low resilience = 11.85 vs. 88.15%).

	E. coli (100mL ⁻¹)	Conductivity (mSm ⁻¹)	Nitrate (mg N·L ⁻¹)	Probability – bad WQ	Odds
Threshold	1000	30	6		
Land use	Return intervals				
Natural	20.87	500	500	5.19	N/A
+Agric	5.32	500	500	19.20	3.7
+Rural	4.39	490	490	23.19	4.5
+Urban	1.97	3.44	12.02	88.15	17.0
+Urban+Wetland	2.03	5.38	22.21	72.35	13.9

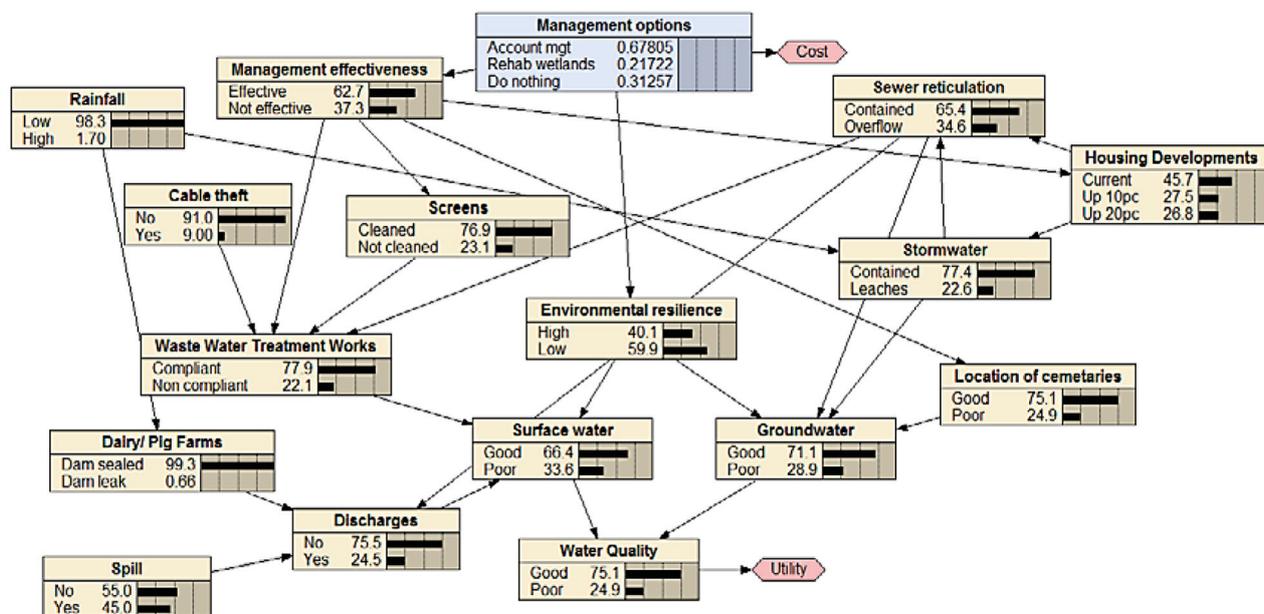


Figure 4
Bayesian network after 3 engagements with the upper Mgeni CMF, and additional research and interaction with individual stakeholders, and incorporating decision and utility nodes

In this example of the BN, under existing prior probabilities (based on best available data), the likelihood that water quality will be 'poor', based on the current set of prior probabilities is 25% (Fig. 4). However, the model still indicates that the utility value of improved management accountability is more than twice the value of doing nothing. Certain variables exhibited considerably more leverage on final 'good' water quality probabilities. For example, changing cable theft from 100% 'no' to 100% 'yes' reduced the probability of good water quality from 75.4% to 72.3%, even though the impacts on likelihood of WWTW compliance were impacted more directly (79.9% compliance probability reduced to 57.7%). Similarly, altering the probability of farm dams from 100% 'sealed' to 100% 'leak' had a marginal impact on final water quality (75.1% probability of good water quality versus 72.5%), while large rainfall events reduced the probability of good water quality from 75.2% to 70.7%. Conversely, other variables had much larger leverage on final water quality probabilities: changes in cumulative housing developments from current to up by 10 and 20% reduced the probabilities of good water quality from 84.5 to 72.0% and 62.3%, respectively. Interestingly, the relative utility of increasing management accountability versus a do-nothing option increased from 1.4 to 1.96 times and 3.52 times for current, 10 and 20% increased cumulative housing developments, such that the value of better management increased as catchment development increased. Similarly, the value of effective management was shown by a reduction in the probability of good water quality from 82.6% to 62.6% when the management effectiveness states changed from 'effective' to 'not effective'.

DISCUSSION

Solutions to problems in aquatic systems often embody conflicting societal interests, and simple models are good because they can be readily adapted to suit needs, and allow for rapid prototyping of ideas that induces systemic thinking on the part of the stakeholders, who are in fact joint model developers.

Water quality issues in South Africa are a typical illustration of complex and conflicting societal interests. The upper Mgeni catchment is no exception, and almost 20 years ago it was already recognised that provision of high-quality water to the main catchment urban centres could only be achieved through holistic catchment management (Little, 1996). As has been illustrated by this study, there are many interacting variables influencing water quality issues, even at the small scale of point-source water quality problems as viewed from a municipal supply perspective. Climate change will only exacerbate these problems, through impacts on rainfall, leading to knock-on effects on changes in runoff, groundwater mineral composition and thus changes in water treatment needs (WHO, 2011).

This study was successful in that many of the issues, previously viewed by the CMF members as not connected, were successfully captured and quantified in a BN, during a series of CMF stakeholder engagement processes. Where this BN has acknowledged weakness is that the conditional probability tables in the network model were populated by the researcher, without following a formal expert elicitation process. This was a consequence of the process only being tested as a pilot approach within a larger research project; funding was thus not available for hosting dedicated workshops for this purpose. Given that all presentations were delivered within the constraints of a quarterly CMF meeting, time was not available for a longer process within the existing meeting framework. It is recognised that for this process to move beyond a theoretical exercise to one that achieved buy-in from the CMF members to the idea of BNs as a useful participatory tool, particular stakeholder groups would need to be engaged with where opinions are kept separate (i.e. people are not allowed to influence each other's thinking), prior to a structured process of combining all opinions for the final CPT. Missing in this exercise was a facilitator to assist with the process of data elicitation within dedicated, funded workshops. In its current form, even assuming that all probability values are spurious, the model serves as a useful tool for stimulating discussion around issues, and by illustrating

the inter-connectedness of multiple variables previously not seen to be linked by members of the catchment management forum. Additional value of this model is that it could be used as a structured compliance assessment tool by the NGOs involved with monitoring, and over time the observations of the variable states could accumulate to such a degree that the model itself has enough data to be validated.

Other issues included the perceived lack of mandate by members of the CMF, where concerns were expressed that the group lacked the leverage to affect regional management, and that certain key stakeholders (for example, municipal managers and agriculture) were not directly represented in the group. What has been learned from this exercise is that while it was relatively easy to sell the idea of using BNs as a participatory tool, for such an exercise to be used as a tool to resolve water quality problems requires the following key elements: wider ranging stakeholder engagement process; dedicated workshops (10–15 people) for different stakeholder groups that follow a formal elicitation process for deriving and populating the BNs; a suitable political champion to support the use of BNs as a participatory reasoning support tool; and, finally, ongoing funding to build up a socially robust case file of observations of the different node states as a means of updating and validating the quantitative component of the BN. However, in spite of these shortfalls, two key elements that emerged for developing a BN were humility and an ability to listen.

A number of limitations to the use of BNs exist, although none of these should preclude using this approach in water resources management. Certain of these limitations are not problematic when BNs are used with other reasoning support system tools, while the remaining limitations are merely constraints to be borne in mind when using BNs. Much data, especially in water management, is continuous (Aguilera et al., 2011), and BNs do not allow for the direct use of continuous data. Rather, there is generally a need to reduce the data into discrete states, which invariably results in data loss (Uusitalo, 2007). States can be of different types: numerical values, intervals, probability distribution or categorical definition. It is also informed by the type of data available and level of model parsimony required. Further shortcomings include that feedback loops are not included (Uusitalo, 2007), and spatial and temporal dynamics are not explicitly handled by BNs. Instead, they may be handled by developing a separate network for each site or time period, which are run separately, or at best, set up in series, where the links in networks are considered as the effect of time. This can add significantly to model complexity (Aguilera et al., 2011). Finally, developing a BN relevant to the problem at hand requires the model developer to have skills in stakeholder consultation and eliciting expert knowledge, i.e., the model is only as good as the data-eliciting and stakeholder engagement processes (Uusitalo, 2007). Therefore potential pitfalls to bear in mind are that there may be difficulty in reaching agreement on the BN structure with experts, and in populating the conditional probability tables based on expert opinion.

Data requirements to develop BNs are as intensive as required by the users. Since model development is relatively quick, investment time is not a major consideration in using this approach. In general, data requirements will be low, and the greatest investment is time in identifying and engaging with stakeholder groups. Developing and using a BN should not be seen in isolation, but rather that BNs are coupled with other modelling approaches, e.g., process-based models to provide input into the models. Thus, for example, spatial GIS layers can be used to derive inputs for BNs (Johnson et al., 2011).

Similarly, outputs from BNs can be applied to raster images, to provide a likelihood surface (Johnson et al., 2011).

The BNs developed in this study represent only the first step in what could become a more complex and iterative process. Developing such a network has been a successful process, and shows promise in integrating issues and stakeholders by first integrating quantitative and qualitative data into an interactive conceptual model of the problem. What has emerged from this process as simple lessons is the importance of listening and observing, engaging, and keeping the approach simple. Community buy-in is critical to the success of any problem-solving process, and what is important is taking the time to tap into the resource of often untapped community wisdom (tacit knowledge). Linking back to the previously-quoted media comment that water quality problems in the Mgeni are due to 'poultry farms, effluent from cattle feedlots and informal settlements without sanitation facilities' (Anon., 2011) – the BNs have illustrated that all of these issues contribute to the problem, but not equally. It has been demonstrated that certain water quality variables are more likely to be the cause of problems than others, and that these should be addressed as the greater priority.

ACKNOWLEDGEMENTS

The following are gratefully thanked for their contributions: South Africa's Water Research Commission (Project K5/1984) for funding. The following individuals and organisations are thanked for their particular contributions: Bonani Madikizela; members of the upper Mgeni CMF; Peter Thomson; Pam Haynes, Penny Rees, Steve Terry. Two reviewers are sincerely thanked for their comments and inputs.

REFERENCES

- AGUILERA PA, FERNÁNDEZ A, FERNÁNDEZ R, RUMÍ R and SALMERÓN A (2011) Bayesian networks in environmental modelling. *Environ. Model. Softw.* **26** 1376–1388. <http://dx.doi.org/10.1016/j.envsoft.2011.06.004>
- AMES DP, NEILSON BT, STEVENS DK and LALL U (2005) Using Bayesian networks to model watershed management decisions: an East Canyon Creek study. *J. Hydroinform.* **7** 267–282.
- ANON. (2011) Health warning for SA rivers. *The Witness/SAPA*, 18 August 2011.
- BACON PJ, CAIN JD and HOWARD DC (2002) Belief network models of land manager decisions and land use change. *J. Environ. Manage.* **65** 1–23. <http://dx.doi.org/10.1006/jema.2001.0507>
- BATCHELOR C and CAIN J (1999) Application of belief networks to water management studies. *Agric. Water Manage.* **40** 51–57. [http://dx.doi.org/10.1016/S0378-3774\(98\)00103-6](http://dx.doi.org/10.1016/S0378-3774(98)00103-6)
- CAIN J (2001) Planning improvements in natural resources management: Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond. Centre for Ecology and Hydrology, Wallingford.
- CAIN J, BATCHELOR C and WAUGHDRAY D (2000) Belief networks: A framework for the participatory development of natural resource management strategies. *Environ. Dev. Sustain.* **1** 123–133. <http://dx.doi.org/10.1023/A:1010033215125>
- CSIR (2010) A CSIR perspective of water in South Africa – 2010. CSIR Report No. CSIR/NRE/PW/IR/2011/0012/A ISBN: 978-0-7988-5595-2. CSIR, Pretoria.
- DICKINSON JL, SHIRK J, BONTER D, BONNEY R, CRAIN RL, MARTIN J, PHILLIPS T and PURCELL K (2012) The current state of citizen science as a tool for ecological research and public engagement. *Front. Ecol. Environ.* **10** 291–297. <http://dx.doi.org/10.1890/110236>
- DWA (DEPARTMENT OF WATER AFFAIRS) (2011) 2011 Green Drop Report, Chapter 6: KwaZulu-Natal, pp. 161–207. URL: <https://www.dwa.gov.za/Documents/GD/GDkzn.pdf> (Accessed 14 December 2011).

- DWAF (DEPARTMENT OF WATER AFFAIRS AND FORESTRY, SOUTH AFRICA) (2005) A drinking water quality framework for South Africa. Department of Water Affairs and Forestry, Pretoria. URL: <http://www.dwaf.gov.za/Documents/Other/DWQM/DWQMFrameworkDec05.pdf> (Accessed 14 December 2011).
- EKZNW (EZEMVELO KWAZULU-NATAL WILDLIFE) (2010) 2005 land cover data. EKZNW, Pietermaritzburg. URL: <http://www.kznwildlife.com> (Accessed 24 January 2012).
- ELLERY W, GRENFELL S, GRENFELL M, JAGANATH C, MALAN H and KOTZE D (2009) A method for assessing cumulative impacts on wetland functions at the catchment or landscape scale. WRC Report No. TT 437/09. Water Research Commission, Pretoria.
- ERŐS T, SCHMERA D and SCHICK RS (2011) Network thinking in riverscape conservation – A graph-based approach. *Biol. Cons.* **144** 184–192. <http://dx.doi.org/10.1016/j.biocon.2010.08.013>
- JENSEN FV and NIELSEN TD (2007) *Bayesian Networks and Decision Graphs* (2nd edn). Springer, New York.
- JOHNSON S, LOW-CHOY S and MENGERSEN K (2012) Integrating Bayesian networks in geographic information systems: good practice examples. *Integ. Environ. Assess. Manage.* **8** 473–479. <http://dx.doi.org/10.1002/ieam.262>
- KALLICHARAN RB (2010) Land use changes within the uMngeni and Mpofana Municipalities: An assessment of environmental impact assessments undertaken during the period 1999 – 2010. Masters thesis, School of Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg.
- KJAERULFF UB and MADSEN AL (2008) *Bayesian Networks and Influence Diagrams: A Guide To Construction and Analysis*. Springer, New York.
- LITTLE AM (1996) A review of the 1970 policy relating to the controlled area of Midmar: Proceeding from the workshop held on 23rd March 1996. Institute of Natural Resources, Pietermaritzburg.
- LYNAM T, DREWRY J, HIGHAM W and MITCHELL C (2010) Adaptive modelling for adaptive water quality management in the Great Barrier Reef region, Australia. *Environ. Model. Softw.* **25** 1291–1301. <http://dx.doi.org/10.1016/j.envsoft.2009.09.013>
- LYNCH SD (2004) Development of a raster database of annual, monthly and daily rainfall for southern Africa. WRC Report No. 1156/1/04. Water Research Commission, Pretoria.
- MANTEL SK, HUGHES DA and MULLER NWJ (2010) Ecological impacts of small dams on South African rivers Part I: Drivers of change – water quantity and quality. *Water SA* **36** 351–360.
- MIDGLEY DC, PITMAN WV and MIDDLETON BJ (1994) Surface Water Resources of South Africa 1990, Volumes I to VI. WRC Report Nos. 298/1.1/94 to 298/6.1/94. Water Research Commission, Pretoria.
- MOLL EJ (1965) An account of the plant ecology of the Upper Mgeni Catchment. MSc thesis, University of Natal, Pietermaritzburg.
- NATHAN R, GORDON N, MCMAHON T, FINLAYSON B and GIPPEL C (2007) AQUAPAK v 1.05. URL: <http://www.skmconsulting.com/Markets/Australia/Water--Environment/Natural-Resource-Management/AQUAPAK-Download.aspx> (Accessed 24 January 2012)
- NILSSON C and RENÖFÄLT BM (2008) Linking flow regime and water quality in rivers: A challenge to adaptive catchment management. *Ecol. Soc.* **13** (2) 18 [online] URL: <http://www.ecologyandsociety.org/vol13/iss2/art18/>.
- NORSYS SOFTWARE CORP (2010) Netica v. 4.16. URL: <http://www.norsys.com> (accessed 13 February 2012)
- RIVERS-MOORE NA (1997) The use of a geographic information system to examine changes in land use patterns in the Midmar Catchment. MSc thesis, University of Natal, Pietermaritzburg.
- SCHMIDT EJ, SCHULZE RE and DENT MC (1987) Flood volume and peak discharge from small catchments in Southern Africa, based on the SCS technique: Appendices. WRC Report No. TT/32/87 and Univ. Natal, Pietermaritzburg, Dept. Agric. Eng., ACRU Report, 26. Water Research Commission, Pretoria.
- SCOTNEY DM (1970) Soils and land use planning in the Howick Extension Area. PhD Thesis, University of Natal, Pietermaritzburg.
- SHENTON W, HART BT and BRODIE J (2010) A Bayesian network model linking nutrient management actions in the Tully catchment (northern Queensland) with Great Barrier Reef condition. *Mar. Freshwater Res.* **61** 587–595. <http://dx.doi.org/10.1071/MF09093>
- STEWART-KOSTER B, BUNN SE, MACKAY SJ, POFF NL, NAIMAN RJ and LAKE PS (2010) The use of Bayesian Networks to guide investments in flow and catchment restoration for impaired river ecosystems. *Freshwater Biol.* **55** 243–260. <http://dx.doi.org/10.1111/j.1365-2427.2009.02219.x>
- TERRY S (2012) Personal communication 25 January 2012. Steve Terry, Scientist, Water Quality and Environment Department, Umgeni Water (Steve.terry@umgeni.co.za).
- UUSITALO L (2007) Advantages and challenges of Bayesian networks in environmental modelling. *Ecol. Model.* **203** 312–318. <http://dx.doi.org/10.1016/j.ecolmodel.2006.11.033>
- WEBB JA, STEWARDSON MJ and KOSTER WM (2010) Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshwater Biol.* **55** 108–126. <http://dx.doi.org/10.1111/j.1365-2427.2009.02205.x>
- WHO (2011) Guidelines for Drinking Water Quality (4th edn). WHO, Geneva. URL: http://www.who.int/water_sanitation_health/publications/2011/dwq_chapters/en/index.html.
- WRC (WATER RESEARCH COMMISSION, SOUTH AFRICA) (2002) State-of-Rivers Report uMngeni River and neighbouring rivers and streams. WRC Report No. TT200/02. Water Research Commission, Pretoria.
- ZORILLA P, CARMONA G, DE LA HERA A, VARELA-ORTEGA C, MARTINEZ-SANTOS P, BROMLEY J and HENRIKSON HJ (2010) Evaluation of Bayesian Networks in Participatory Water Resources Management, Upper Guadiana Basin, Spain. *Ecol. Soc.* **15** 12 [online] URL: <http://www.ecologyandsociety.org/vol15/iss3/art12/>.