The use of Bayesian Belief networks to aid in the understanding and management of large-scale coral bleaching

Scott Wooldridge and **Terry Done**

Reef Futures Project, Australian Institute of Marine Science, Townsville, Australia

Abstract: In this paper we utilise a collection of modelling tools based on the Bayesian paradigm, to develop a framework that is capable of delivering the continuum of "value added" science; the progression from the acquisition of field-based and remotely-sensed evidence, through to the development of decision support structures that have the potential to assist decision making for the management of complex systems. By way of example, we utilize the approach to identify the relative strengths of dependency of a range of potential causative factors associated with the presence-absence of large-scale coral bleaching. For the factors considered, we show that short-term maximums in sea surface temperatures provide the strongest causative link to bleaching related coral mortality. We demonstrate however that the composition of the coral community affords some level of protection against temperature related mortality. We discuss the benefits of the outlined methodology in terms of providing a decision support platform to aid in the management of large-scale coral reefs ecosystems; a system in which current and future uncertainties can not be ignored.

Keywords: *Bayesian Belief networks; decision support; coral bleaching; large-scale management*

1. INTRODUCTION

Traditional coral reef science has been extremely successful in documenting the infinite complexity of coral ecosystems. It has been able to demonstrate that this complexity arises from the fact that coral ecosystems are richly connected via processes and feedback mechanisms that operate over a variety of spatial and temporal scales. Beyond the appreciation for this complexity, the large-scale management of coral reef ecosystems is now heavily reliant on the ability to forecast the potential impacts of change (e.g. due to climate change or any other form of human impact), such that appropriate preventative measures can be implemented.

It is clear, that at any level of complexity, simulation models for the management of coral reef systems will never replicate the complexity in nature. Less clear is the conclusion that a prudent environmental modelling strategy is to avoid detailed mathematical characterization of natural processes for models developed to assist coral reef policy-makers and managers. This recommendation, from those who have studied environmental simulation and policy (e.g. Hodges, 1987) is based on the observation that the most useful predictive management models are often extremely simple, or at least conceptually simple.

The questions of interest to decision makers broadly concern the relationship between a management option and an attribute of concern to the public e.g. "over what area and time-frame can we expect improvements in the growth and recovery of inshore coral reefs as new water quality targets are met?" Complex environmental simulation models are not essential to answer questions of this nature, indeed there is to much unpredictability in nature to mathematically describe all the mechanisms affecting coral growth in natural waters. An alternative approach is to *embrace* the inherent uncertainty in natural systems. This method has been employed by the physicist who uses probabilistic expressions to capture the aggregate response of molecular motion in statistical mechanics. In a similar manner, a coral ecologist could summarise small-scale processes with probabilistic expressions that characterize the aggregate response of interest to the decision maker e.g. application of a useful predictive framework might yield the following statement: "if annual nutrient loading to river X is reduced by 40% or more, the decadal probability of major coral mortality is less than 0.2".

These observations about models and prediction reflect a decision analytic perspective (Clemen, 1996) and lead to the consideration of influence diagrams and Bayesian Belief network (BBN) models (Pearl, 1988). A BBN is a 'causal reasoning' tool that has recently attracted an increasing number of researchers in the field of applied artificial intelligence where uncertainty is an intrinsic characteristic of the problem domain. The Bayesian and conditional probability theory upon which BBNs are founded, provides a mathematical framework for updating one's belief in the occurrence of an outcome or event, given the observation of certain pieces of evidence;

emulating the way in which and expert might be expected to make decisions within an uncertain environment.

In this paper, we demonstrate the application of the Bayesian paradigm, and in particular the decision support capabilities provided by BBNs for the task of identifying the relative importance (i.e. dependency) of a range of potential causative factors associated with the presence-absence of large-scale coral bleaching. We approach the problem from the standpoint of letting the 'information content' of the data tell the story about the level of complexity and relative strengths of any dependencies. As such, we endeavour to integrate spatial data sets from a variety of sources, including satellite-derived sea surface temperature (SST) measurements, ocean current predictions from a physical oceanography model, and bathymetry data from a digital elevation model.

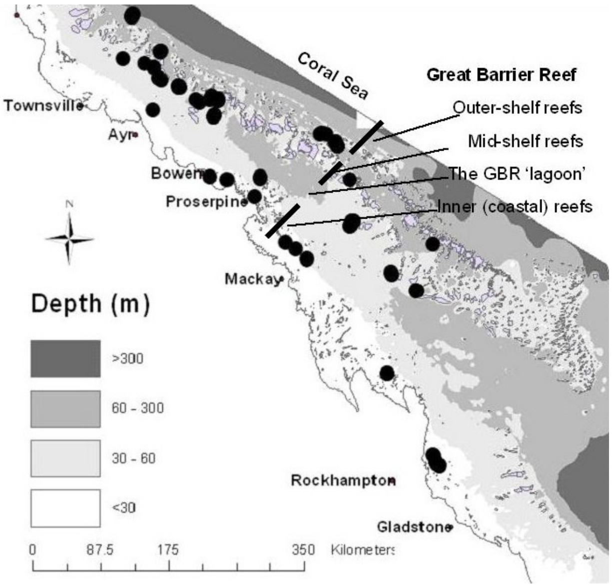

Figure 1. Survey sites within the GBR study area.

2. CASE STUDY: IDENTIFYING CORAL REEFS WITH A LOW *RISK-TO-BLEACHING*

2.1. Background

The term 'coral bleaching' summarises one of the most insidious large-scale hazards faced by coral reefs: environmentally stressed corals become visually pale (or bleach white) through (i) the expulsion of damaged zooxanthellae or (ii) the reduction in photosynthetic pigments within the zooxanthellae, or (iii) a combination of both responses (see e.g., Berkelmans, 2002). Coral bleaching is a generalised stress response caused by prolonged exposure to anomalous environmental conditions; in particular, sea temperatures that are more than 1°C higher than the long-term mean summer maximum for the location in which the coral resides. Continued expose to anomalous conditions often results in the bleached coral experiencing partial or whole colony mortality. Alternatively, the return to 'normal' environmental conditions within specific time-frames (Berkelmans, 2002) can result in the apparent recovery to pre-bleached conditions.

During the southern-hemisphere summers of 1998 and 2002, large regions of the Great Barrier Reef (GBR) experienced unprecedented heating anomalies and associated coral bleaching/mortality. Climate models suggest a future with increased frequency of these anomalous events, so there is an imperative to identify low *risk-to-bleaching* areas; which could form logical sub-units in the design of bleaching-resistant marine protected areas with the potential to be reservoirs of abundance and biodiversity in coming decades.

2.2. The 2002 GBR Bleaching Data Sets

Ecological characterization and bleaching impact assessments

Following the 2002 bleaching event, a series of three field cruises were undertaken over a total of 30 days between June 25 and August 1. The surveys assessed the environmental setting, bleaching impact and coral mortality for coral communities from 150 sites at 50 locations on 32 reefs. The survey locations ranged from the southern to the central GBR (~1000 km), and from less than 1 km from the coast to >200 km out into the Coral Sea (Fig. 1).

Each site was assigned a habitat class. For the mid- and outer shelf reefs, four classes were identified; outer slope, lagoon, back-reef, or channel. For the inner, coastal reefs, a single 'fringing reef' class was used. Assessments involved a 20-30 minute survey, during which a taxonomic inventory was complied of observed hard and soft corals, and a bleaching impact category noted for each taxon.

Indicators of bleaching and coral mortality were developed from the visual assessment data, and four 'community types' were developed using K-mean clustering (see Done et al., 2003 for details). Types 1, 2 and 3 were offshore (mid- and outer shelf) reefs, whereas Type 4 communities were coastal.

Heat Stress Characterisation of the thermal environment

Maximum Heat Stress: Following suggestions from Berkelmans et al. (2003) and using 1km^2 SST estimates as derived from advanced very

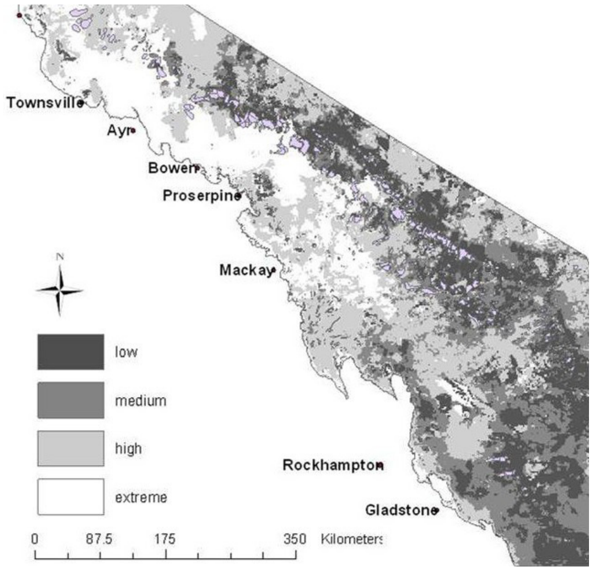

Figure 2. *Maximum heat stress* indicator

high resolution radiometer (AVHRR) sensors aboard the NOAA16 satellite, we developed a proxy *heat stress* indictor for each 1km^2 pixel based on the highest accumulated SST total for any three-day run of summer SST (2002). Figure 2 displays the spatial variation of the 3-day *heat stress* indicator for the survey region. The map indicates that very hot water bathed many reefs in the survey region, right across the reef tract from the coast to Coral Sea. On the other hand, it also indicates how much of the outer reef tract escaped exposure to even brief periods of hot water.

Patterns of Cooling: Waters below the thermocline are important sources of cooling for shallow water corals if they can be mixed into surface waters. As a proxy indicator for the 'ease' of access of shallow corals to this cool oceanic water, we used hydrodynamic model predictions of *average* tidal current strength to derive a GIS-based cost surface that indicates the 'effective' distance for all points from the 100 m isobath which borders the GBR and Coral Sea (Fig. 3). To generate the cost surface, the range of currents within the study region were normalised between 0 and 1; the lower the number the larger the strength of currents (i.e. lower flow resistance). In traveling from the 100m isobath, the linear sum of all 1 km^2 pixels encountered in reaching any point constituted its effective 'cost'. Since the dominant water movement from the prevailing seas is from the south-east, the linear sum was calculated for a north-westerly direction of travel. The final 'accumulated' cost surface was then also normalised to lie in the range 0 to 1. The general hypothesis is that the greater the cost, the lower the cooling potential from this source.

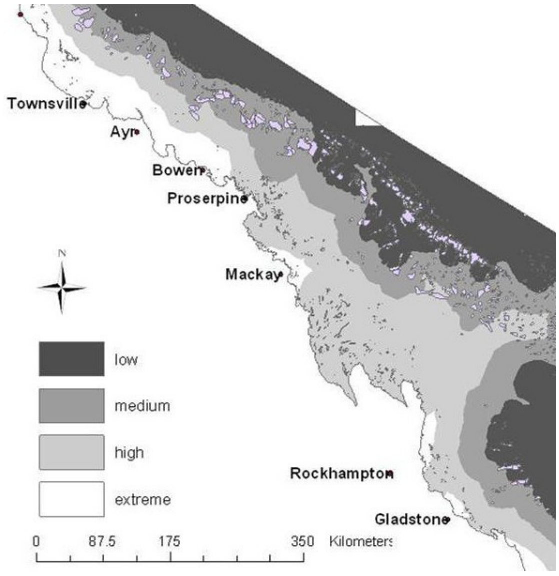

Figure 3. *Patterns of Cooling*, indicating the 'effective' distance from the 100m isobath.

2.3. Results

We now have a series of spatial indicators to describe the potential thermal environment of our study sites, along with the field-derived proxies for bleaching impacts, coral mortality, and coral assemblages (i.e. community type). The challenge now is to see the extent to which the 'information content' of our proxies allows us to predict locations vulnerable to bleaching and coral mortality; in particular the ability to infer low *risk-to-bleaching* areas.

An exploratory Bayesian analysis

We adopted novel artificial intelligence methods to explore our proxy data-sets in an integrated way. Initially, we used the *dependency analysis* algorithm of Cheng et al. (2002), to search for structural dependency relations among the information contained within our proxy indicators. For ease of reference, Table 1 provides the name and brief description of the indicator variables used in the analysis.

To aid the *dependency analysis* process, we imposed the following 'expert' domain knowledge (based on the current beliefs about bleaching resistance, and refined in this study):

- *cost100* and *habitat* are externally determined drivers that cannot be affected by anything else in the coral reef system,
- *community* has a direct dependency on *habitat*, and
- *bleach* and *dead* are hypothesis (i.e. event) variables, which can potentially accept dependency linkages from all system variables.

Table 1. Summary of the proxy variables used in the *dependency analysis.*

Variable	Description
cost100	'Effective' distance from 100 m depth
habitat	Classified habitat class
community	Classified coral community types
max3day	Max. 3-day run of summer SST (2002)
bleach	Classified bleaching impact
dead	Classified coral mortality

Bayesian Belief network development

The identified dependency structure, along with the associated strength of the linkages, was used to construct a BBN. The BBN encodes the dependency relationships between our proxy indicator variables; which are represented as *nodes*. Individual nodes are constrained to contain the finite number of mutually exclusive states (e.g. high, medium, low) that describe our proxy variables. Nodes are connected by *arcs* (dependency links), which point from *parent* nodes (causes) to *child* nodes (effects). The absence of a link between two variables indicates independence between them. The strength (i.e. certainty) of the causal link between a *child* and its *parent* node(s) is summarised through a conditional probability distribution usually in the form of a conditional probability table (CPT). The CPT specifies the conditional probability of the child node being in a particular state, given the states of all its parents: $P(child \mid parent_1, parent_2, \ldots parent_N)$. Should a node have no parents, the table reduces to an unconditional one: $P(child)$. Given the structure of the BBN, and the associated conditional probabilities, it is possible to determine the likelihoods of different states in each *child*, given the likelihoods of different states in its *parent(s).* The power of the BBN comes to light whenever we change the likelihood of parent states, based on field evidence or expert opinion. The effects of the evidence or opinion are propagated throughout the dependence-structured network via a 'probabilistic inference algorithm', and the resulting probabilities of the affected nodes updated (see Lauritzen and Spiegelhalter, 1988).

We used the *Netica* (http://www.norsys.com) network editor to depict the dependency structure identified from our available domain variables (Fig. 4). Graphically, it can be seen that as expected, the state of the heat stress variable *max3day*, was found to have a dependency on *cost100*, while it in turn provides a level of dependency for the states of the *bleach* and *dead* nodal variables. Along with the enforced

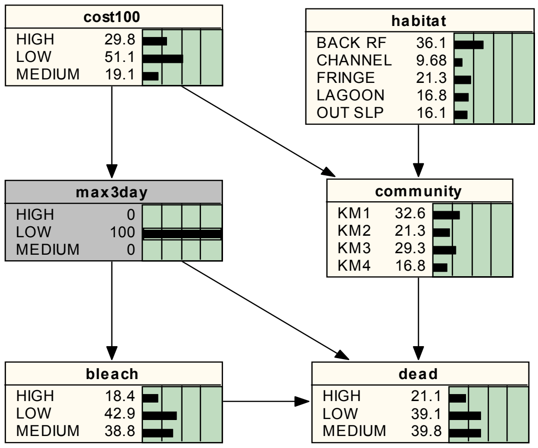

Figure 4. BBN1 – predictive model of coral mortality, which includes 'evidence' nodes to describe both the thermal environment and coral reef characteristics.

dependency on *habitat*, the *community* nodal variable was also shown to have some level of dependency on *cost100*; potentially reflecting how GBR coral community types 'track' water motion and wave impact gradients (Done, 1982). No link was established between *community* and *bleach*, but a direct linkage was found between the *community* and *dead* nodal variables; perhaps reflecting that the information content of the *bleach* variable is a little confounded due to the lateness of the survey (by the time of the June to August surveys, some corals that bleached in February would have regained normal colour).

Rather than providing descriptions of the CPTs that indicate the strength (i.e. certainty) of the developed dependency linkages, we instead chose to undertake a sensitivity analysis to identify the network components which have the greatest influence on coral mortality. The sensitivity analysis was conducted by systematically varying the values of individual network components to determine how they affected the *dead* nodal variable; the results highlighting how much the mean belief value of the *dead* node could be influenced by a single finding at each of the other nodes in the network. We restricted our summary to just the *high* state. From Figure 5 we can see that for the *high* state of *dead*, the *bleach* and *max3day* nodal variables are the most influential components within the network. Due to the potential for high correlation between the level of bleaching and the level of mortality we would expect that knowledge of *bleach* should provide a high level of inference on *dead*. Of more interest, is the relative importance of *max3day* compared to the other nodal variables. The result suggests that in the absence of any other information, selection of sites in areas classified as *low* for the *max3day* indicator would greatly improve the

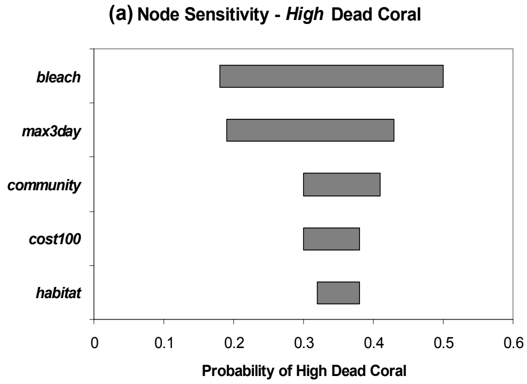

Figure 5. Sensitivity of *high* dead coral cover to changes in the nodes of BBN1.

likelihood of including low impact sites. Of the other nodal variables, *community* provides the next highest level of influence, but it is considerably less influential (based on whole system performance) than the thermal stress information.

Bayesian Belief network predictions

The value of the developed BBN as a predictive decision support tool was also tested. We used the *Netica* runtime algorithm (Lauritzen and Spiegelhalter, 1988) to propagate the available evidence at each site. As shown in Table 2a, the identified BBN correctly assigns 106/150 sites (predictive rate 71%) to their actual observed coral mortality class – this compares favorably with 44% by "informed guessing", i.e. guessing all reefs had *medium* mortality, and 33% by "blind guessing". Essentially equal predictive ability was achieved for *low*, *medium* and *high* levels of coral mortality.

To test the predictive value of including the *community* node information, beyond that contained by the *max3day* node alone, we developed a new BBN in which the *dead* variable had no dependency on *community* (Fig. 6). The strengths of the remaining conditional linkages were then re-learnt from the original database. Surprisingly, as shown in Table 2b, for the new BBN, only 91/150 sites (predictive rate 61%) were correctly assigned. Interestingly, the majority of the loss in predictive capacity can be apportioned to the inability to correctly predict the *low* coral mortality class (36% compared with 72% when the *community* node was included). This is significant, given that our ultimate objective is to correctly identify areas with *low* coral mortality potential.

With this objective in mind, we compared the predictive ability of the original (BBN1 – Fig. 4) and modified (BBN2 – Fig. 6) networks, in

Table 2: Actual (i.e. observed) and predicted coral mortality for the 150 survey sites.

	Predicted Coral Mortality			Actual	Predictive Rate
(a)	*Low*	*Med.*	*High*		
BBN1	**26**	7	3	*Low*	*26/36=0.72*
	11	**46**	9	*Med.*	*46/66=0.70*
	1	13	**34**	*High*	*34/48=0.71*
(b)	*Low*	*Med.*	*High*		
BBN2	**13**	18	5	*Low*	*13/36=0.36*
	6	**46**	14	*Med.*	*46/66=0.70*
	1	15	**32**	*High*	*32/48=0.67*

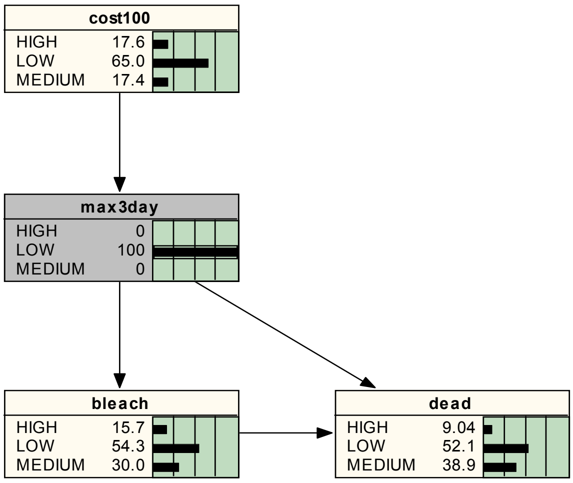

Figure 6. BBN2 –predictive model of coral mortality, which includes 'evidence' nodes to describe only the thermal environment.

identifying those sites which were observed to have *low* coral mortality, despite being in warm-high heat stress conditions. Interestingly, of the 15 occasions when *low* mortality was observed in conjunction with *medium-high* heat stress conditions, BBN1 was able to correctly predict 9 of them. On the other hand BBN2 was unable to correctly predict any of them. The result clearly highlights the potential for differential mortality responses to heating stress based on the resident community type. Demonstrating that many of the subtleties within the system are conditional responses; a property that is efficiently captured by BBN1.

On further inspection, community type 3 (i.e. *KM3*) was found to be dominate when BBN1 correctly predicted *low* mortality given *medium-high* heat stress conditions. *KM3* was a low diversity *Acorpora/Faviid* community of reef flats and shallow slopes. The result suggests that inter-tidal or shallow sub-tidal reef flat sites - by virtue of shallow depth and periodic exposure to air during low tides – may be relatively more sun-hardened and/or heat-hardened than reef slope sites. This result deserves further inspection and testing, as it suggests that community level adaptation (i.e. the favouring of certain suites of species) may be a potential adaptive response to future sea warming.

This study is based on a relatively small survey sample. However, it is our belief that the 'information content' of our ecological data sets has the potential to be strong due to the fact that the study locations were not just randomly selected but benefited from a process of informed selection (guided by satellite SST data) thereby targeting the thermal conditions of interest, and limiting redundancy in the data-set. Secondly, the employed *dependency analysis* algorithm (Cheng et al., 2002), benefited from being able to accept 'expert' guidance on the initial structure; freeing the information content of the data to aid in the discovery of weaker dependency signals. Regardless of these facts, as with any modeling exercise, the work presented would benefit from evaluation against an independent data set. This is the subject of on-going research and will be reported elsewhere.

3. DISCUSSION AND CONCLUSIONS

In this paper, we have demonstrated how the Bayesian paradigm, and in particular the decision support capabilities provided by BBNs can help to bridge the growing disconnect between the outputs currently delivered by coral reef science and the requirements of management. BBNs were shown to be able to accept all the essential elements of 'understanding' – data, information, and knowledge (even in the form of expert opinion) – and fuse them into a framework that is capable of delivering decision support in the form of prognostic evaluations of the likelihood of achieving specific outcomes or *event* states (in our case, bleaching and mortality status).

This fusion capability provides an excellent vehicle for an interdisciplinary focus to problems, indeed in the development of our BBNs, we made use of the knowledge and information provided by ecologists, oceanographers, climatologists, bio-statisticians, and spatial analysts. Importantly, because uncertainty in particular inter-linkages is acknowledged in the conditional probabilistic statement of relationships, the decision support capability of the developed BBNs were not limited by the need to account for all mechanistic detail. Indeed, by incorporating the 'experience of past events' as well as developed spatial relations, we allowed the level of descriptive complexity (i.e. dependency structure) to be 'learnt' from the information content of the available data. The flexibility provided by BBNs also means that as new information becomes available, it is a trivial task for it to be incorporated, with only the conditional probabilities of the affected variables requiring re-evaluation. For complex adaptive systems like coral ecosystems, this is important since we are continually going to experience 'surprises' in the magnitude (and potential type) of stress and recovery responses. But unlike methods that require us to continually throw away and start with a clean sheet, the BBN learning process allows us to adapt to those surprises or mistakes, which ultimately, we must expect to be our most valuable sources of information and experience.

4. ACKNOWLEDGEMENTS

The authors wish to thank R. Berkelmans, G. De'ath, S. Edgar, S. Kininmonth, M. Mahoney, C. Steinberg, E. Turak, and M. Wakeford, all from the Australian Institute of Marine Science, for their help in various aspects of this study.

5. REFERENCES

Berkelmans, R., Time-integrated thermal bleaching thresholds of reefs and their variation on the GBR, *Mar Ecol Prog Ser*, 229, 73-82, 2002.

Berkelmans, R., De'ath, D., Kininmonth, S., and W., Skirving, Coral bleaching on the GBR: correlation with SST, and comparison of the 1998 and 2002 event, *Coral Reefs*, in press, 2003.

Cheng, J., Griener, R., Kelly, J., Bell, D., and W., Liu, Learning Bayesian networks from data: An information-theory based approach, *Artificial Intelligence*, 137, 43-90, 2002

Clemen, R.T., Making hard decisions, Duxbury Press, Belmont California, 1996.

Done, T., Patterns in the distribution of coral communities across the central GBR, *Coral Reefs*, 1, 75-90, 1982.

Done, T., and others, Testing bleaching resistance hypotheses for the 2002 GBR bleaching event. Published report prepared for The Nature Conservancy by the Australian Institute of Marine Science, 2003

Hodges, J.S., Uncertainty, policy analysis, and statistics, *J. Am. Stat. Assoc*., 2, 259-291, 1987.

Lauritzen, S., and D. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, in *J. Royal Statistics Society B*, 50(2), 157-194, 1988.

Pearl, J., Probabilistic reasoning in intelligent systems: Networks of plausible inference, Morgan Kaufmann Publishers, San Francisco, Calif, 1988.