18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

A Bayesian Network approach to modelling temporal behaviour of *Lyngbya majuscula* bloom initiation

Johnson, S. [1] **and K. Mengersen** [1]

[1] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland*
Email: s6.johnson@qut.edu.au

Abstract: *Lyngbya majuscula* is a cyanobacterium (blue-green algae) occurring naturally in tropical and subtropical coastal areas worldwide. Deception Bay, in Northern Moreton Bay, Queensland, has a history of *Lyngbya* blooms, and forms a case study for this investigation.

The South East Queensland (SEQ) Healthy Waterways Partnership, collaboration between government, industry, research and the community, was formed to address issues affecting the health of the river catchments and waterways of South East Queensland. The Partnership coordinated the *Lyngbya* Research and Management Program (2005-2007) which culminated in a Coastal Algal Blooms (CAB) Action Plan for harmful and nuisance algal blooms, such as *Lyngbya majuscula*. This first phase of the project was predominantly of a scientific nature and also facilitated the collection of additional data to better understand *Lyngbya* blooms. The second phase of this project, SEQ Healthy Waterways Strategy 2007-2012, is now underway to implement the CAB Action Plan and as such is more management focussed.

As part of the first phase of the project, a Science model for the initiation of a *Lyngbya* bloom was built using Bayesian Networks (BN). The structure of the Science Bayesian Network was built by the *Lyngbya* Science Working Group (LSWG) which was drawn from diverse disciplines. The BN was then quantified with annual data and expert knowledge. Scenario testing confirmed the expected temporal nature of bloom initiation and it was recommended that the next version of the BN be extended to take this into account. Elicitation for this BN thus occurred at three levels: design, quantification and verification. The first level involved construction of the conceptual model itself, definition of the nodes within the model and identification of sources of information to quantify the nodes. The second level included elicitation of expert opinion and representation of this information in a form suitable for inclusion in the BN. The third and final level concerned the specification of scenarios used to verify the model.

The second phase of the project provides the opportunity to update the network with the newly collected detailed data obtained during the previous phase of the project. Specifically the temporal nature of *Lyngbya* blooms is of interest. Management efforts need to be directed to the most vulnerable periods to bloom initiation in the Bay. To model the temporal aspects of *Lyngbya* we are using Object Oriented Bayesian networks (OOBN) to create 'time slices' for each of the periods of interest during the summer. OOBNs provide a framework to simplify knowledge representation and facilitate reuse of nodes and network fragments. An OOBN is more hierarchical than a traditional BN with any sub-network able to contain other sub-networks. Connectivity between OOBNs is an important feature and allows information flow between the time slices.

This study demonstrates more sophisticated use of expert information within Bayesian networks, which combine expert knowledge with data (categorized using expert-defined thresholds) within an expert-defined model structure. Based on the results from the verification process the experts are able to target areas requiring greater precision and those exhibiting temporal behaviour. The time slices incorporate the data for that time period for each of the temporal nodes (instead of using the annual data from the previous static Science BN) and include lag effects to allow the effect from one time slice to flow to the next time slice. We demonstrate a concurrent steady increase in the probability of initiation of a *Lyngbya* bloom and conclude that the inclusion of temporal aspects in the BN model is consistent with the perceptions of *Lyngbya* behaviour held by the stakeholders. This extended model provides a more accurate representation of the increased risk of algal blooms in the summer months and show that the opinions elicited to inform a static BN can be readily extended to a dynamic OOBN, providing more comprehensive information for decision makers.

Keywords: *Algal bloom, Lyngbya majuscula, Bayesian Belief Network, BN, Object Oriented, OOBN*

1. INTRODUCTION

The presence and intensity of *Lyngbya* blooms have been officially recorded by the EPA since 2000 (Hodgkinson and Cox, 2005). Figure 1, which is based on this data, shows a box plot of the monthly intensity of *Lyngbya* blooms in Deception Bay from January 2000 to December 2005. From the box plot it is apparent that the initiation of a *Lyngbya* bloom is of increasing concern with the onset of summer. For this reason it was recommended at the annual *Lyngbya* Science BN Review meeting (1 December 2008) to more accurately represent the temporal nature of *Lyngbya* blooms. The original structure of the BN was elicited from the experts in the Lyngbya Science Working Group during the first phase of the *Lyngbya* Research and Management Program (Hamilton *et al.*, 2007b). The BN structure was also reviewed at the annual BN Review meeting and the resulting modified model is shown in Figure 2. The temporal nodes of interest were identified as *Temperature*, *Rainfall*, *No of Previous Dry Days*, *Surface Light*, *Wind Speed*, and *Wind Direction*. These nodes are depicted in the model below as 'input' nodes (shown as an eclipse with a broken line) which means that they act as place holders for the actual data for the months of September to March.

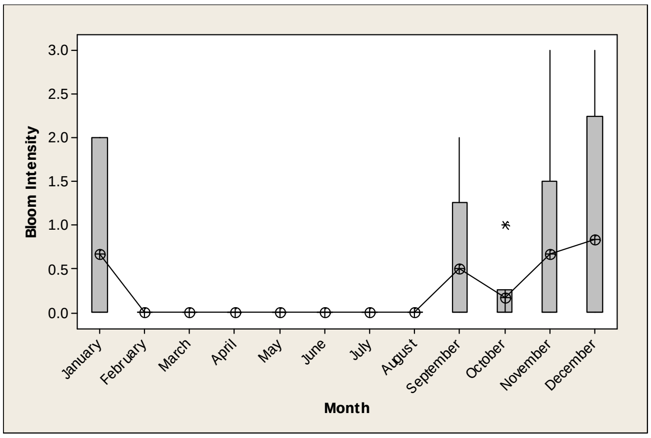

Figure 1. Box plot of intensity of Lyngbya blooms in Deception Bay from January 2000 to December 2005

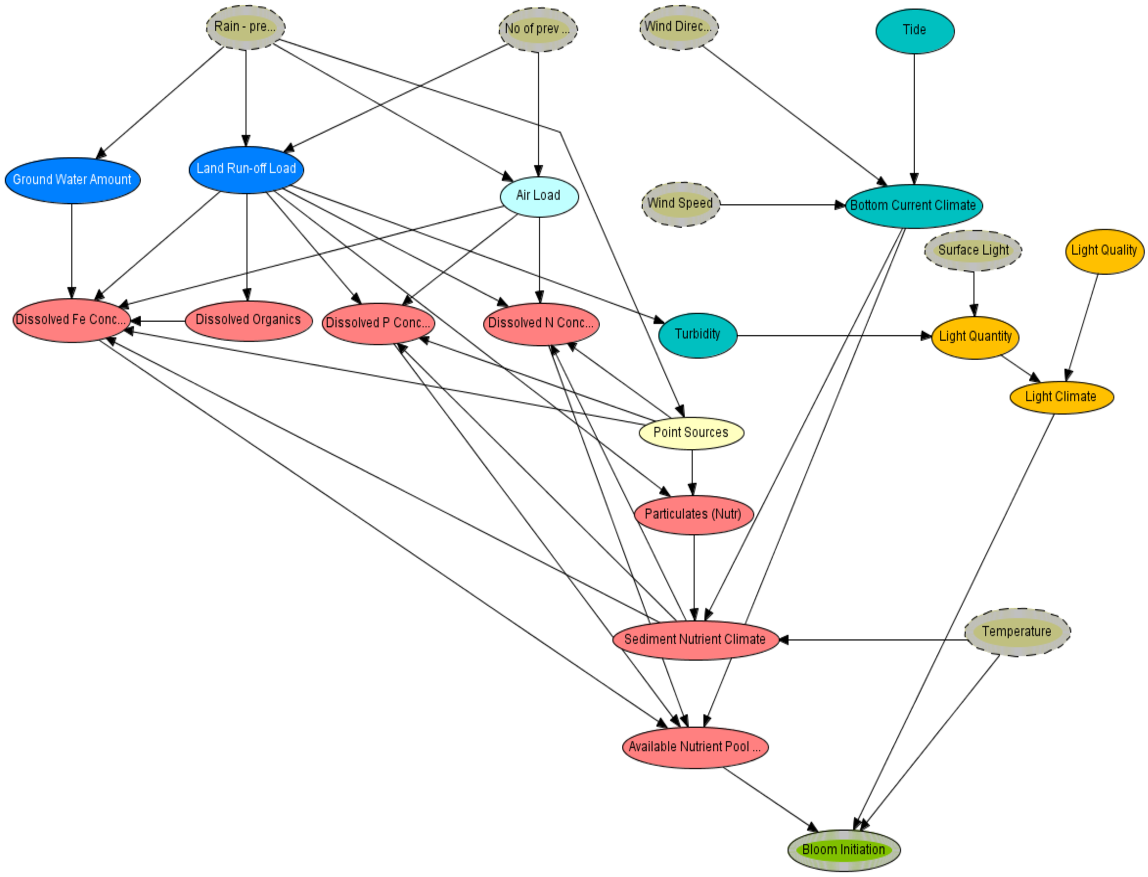

Figure 2. *Lyngbya* Science BN as a generic OOBN showing the input nodes with broken lines (Hugin®)

Hugin® software is popular to model BNs, easy to use and has good object oriented modelling capabilities. Figure 2 shows the *Lyngbya* Science BN as an Object Oriented Bayesian Network (OOBN) which enables it to be used as a 'blue print' for the seven months of interest. In other words it is a generic network which can be used (instantiated via an instance node) in each of the monthly networks. The monthly data is then added for each of the six input nodes to create the OOBN specific for that month.

2. TEMPORAL NODE QUANTIFICATION

Every node in a BN has a conditional probability table (CPT) associated with it. The temporal nodes considered here were populated using the DPI Forestry's Meteorological Records at Beerburrum containing daily readings from November 1999 to November 2005.

2.1. Rainfall

The *Rainfall* node has three expert-defined states: *Low* (0-5mm/day), *Medium* (6-25mm/day) and *High* (25+mm/day). The resulting CPT table which was used to quantify the *Rainfall* node for each of the seven monthly BNs is shown in Table 1. It gives the number of days which were recorded as having low, medium or high rainfall as a percentage of the total number of days for that month (averaged over six years of data). Figure 3 clearly shows that the average number of days with low rainfall is greater than medium or high rainfall days throughout the year. However in the seven months of interest to *Lyngbya* bloom initiation, and also in April, the average number of days having medium and high rainfall is more than in other months. From the *Lyngbya* Science OOBN in Figure 2 we can see that this in turn would have an effect on the *Groundwater Amount* (water stored below the earth's surface), *Land Run-off* (overland flow of water) and *Air Load* (nutrient load from Aeolian sources).

Table 1. Monthly Rainfall (%)

	Low	Medium	High
Sep	97.2	2.8	0.0
Oct	86.0	10.8	3.2
Nov	81.7	12.2	6.1
Dec	77.4	13.4	9.1
Jan	86.0	10.8	3.2
Feb	81.2	10.0	8.8
Mar	81.2	15.6	3.2

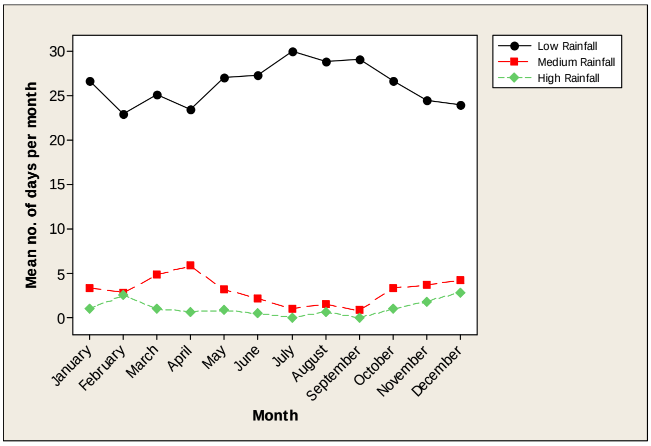

Figure 3. Mean Rainfall patterns (Nov 1999 to Oct 2005)

2.2. Number of Previous Dry Days

Here expert opinion defined the node, *No of prev dry days*, as the cumulative number of days where rainfall is less than 5mm. Expert opinion also determined the states and thresholds for this node as follows: For any day in the month, if the number of previous dry days was 2 or less, it was assigned a value of *Low*; *Medium* indicates 3 to 5 previous dry days and *High* more than 6 days having less than 5mm rain. The CPT to populate this node for each of the seven months is given in Table 2. Figure 4 illustrates the combination of mean number of days with low, medium and high previous dry days for every month. As expected, since Deception Bay is in a summer rainfall area, the graph shows the winter months of June to August having on average the most number of days with a high number of previous dry days. For the seven months being modelled for *Lyngbya* bloom initiation, however, there tend to be a more even distribution between the number of days with low, medium and high previous dry days.

Table 2. Previous number of dry days (%)

	Low	Medium	High
Sep	8.9	8.9	82.2
Oct	31.2	17.2	51.6
Nov	38.9	24.4	36.7
Dec	43.0	19.4	37.6
Jan	30.1	19.4	50.5
Feb	38.8	19.4	41.8
Mar	44.1	23.7	32.3

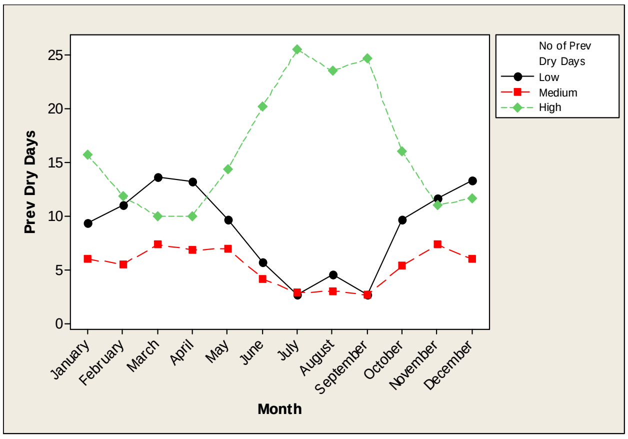

Figure 4. Mean number of days which had low, medium and high previous dry days (Nov 1999 to Oct 2005)

2.3. Temperature

The LWSG defined the *Temperature* node to be the temperature of the water column. It has only two expert-defined states, *Low* or *High*. A study by Hamilton *et al.*, (2007a) evaluating several alternative models, found that a model including only the minimum monthly temperature had the best predictive behaviour for *Lyngbya* bloom initiation. Furthermore air temperature was used as an acceptable approximation for water temperature, only differing typically by approximately one degree. Therefore the CPT for this node (Table 3) used the daily minimum air temperature recorded at Beerburrum. Expert opinion set 17°C as the cut-off point, so that if the minimum daily temperature was more than 17°C it was recorded as a high minimum temperature, otherwise it was classed as low. This cut-off can be varied at a subsequent annual review meeting if there are scientific reasons to differentiate between the states relative to another temperature.

Table 3. CPT for Minimum Temperature (%)

	Low	High
Sep	98.3	1.7
Oct	86.6	13.4
Nov	72.7	27.3
Dec	34.8	65.2
Jan	18.8	81.2
Feb	22.9	77.1
Mar	25.8	74.2

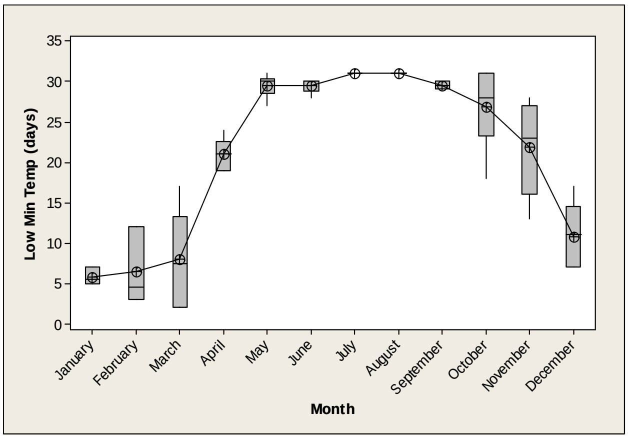

Figure 5. Box plot of the number of days with a low minimum temperature (Jan 2000 to Oct 2005)

Figure 5 is a box plot of the daily minimum air temperatures grouped by month for the six years. The means and medians are generally approximately equal, but the observations appear heavily skewed to the right in February and somewhat skewed in October and November. Nonetheless there appears to be a definite trend for the number of days having low minimum temperatures to start increasing from March and then more rapidly to peak in the winter months, before slowly declining from September and more quickly from October.

2.4. Wind Direction

The *Wind Direction* node was defined by the LSWG to represent the measured course of the wind, relative to the compass. Expert opinion nominated only two states to be relevant to *Lyngbya* bloom initiation: southeast (*SE*) and north (*N*). All other directions were grouped as *Other*. The wind direction at Beerburrum is recorded three times daily (9am, noon or 3pm). Watkinson *et al.* (2005) noted that during *Lyngbya* bloom initiation there were north to northeast (NE) winds. Therefore based on expert guidance, the daily wind direction was designated as N if any of the recorded wind directions for the day was N, NNE or NE. In a similar vein if any reading during the day recorded a SE wind, the daily value was taken to be SE. Figure 6 shows the composition of the mean number of days per month for wind directions of N, SE and Other, and Table 4 shows the percentage allocations for each of the three directions for the months of September to March. Watkinson *et al.* (2005) took wind direction and wind speed from observations made at noon in contrast to the way it was determined here. The flexibility exists to alter the program code if the next review recommends the use of only the noon reading.

Table 4. CPT for Wind Direction (%)

	North	SE	Other
Sep	41.7	28.9	29.4
Oct	52.7	28.5	18.8
Nov	43.3	41.1	15.6
Dec	43.0	35.5	21.5
Jan	30.6	53.8	15.6
Feb	20.0	51.8	28.3
Mar	18.3	53.8	28.0

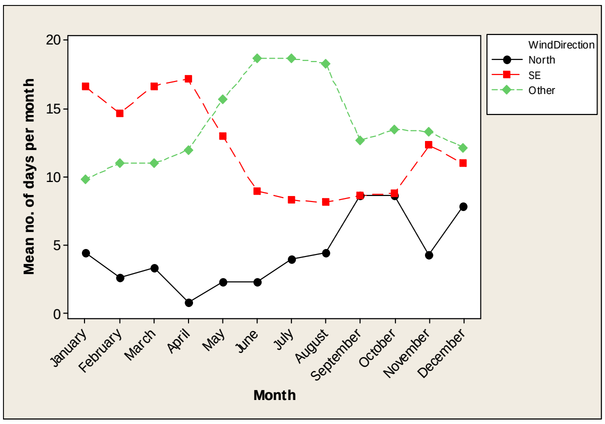

Figure 6. Mean number of days per month for N, SE and Other wind direction (Nov 1999 to Oct 2005)

2.5. Wind Speed

The *Wind Speed* node represents the rate at which the wind travels over the surface of the water and has only two expert-defined states, *High* and *Low*. Using the Beaufort wind force scale, the wind speed reading for the day was assigned a value of *Low* if the average wind force was less than 3 and *High* if it was 3 or more. Table 5 shows the resulting CPT table for this node showing the percentage number of days in that month that were deemed to have a high or low wind speed, based on records from November 1999 to October 2005. The daily average wind speed was calculated using all the daily recorded readings (9am, noon and 3pm) if they were present. This can be changed to the noon reading only, as was done by Watkinson *et al.* (2005). However the average speed appeared to be a good representation of the wind speed for the day and furthermore enables us to use more days of data when the noon reading was missing, but the 9am or 3pm reading was available. Figure 7 shows the box plot of the daily readings which were classified as having high wind speed, grouped by month.

Table 5. CPT for Wind Speed (%)

	Low	High
Sep	80.3	19.7
Oct	91.9	8.1
Nov	89.4	10.6
Dec	91.8	8.2
Jan	87.5	12.5
Feb	95.3	4.7
Mar	91.4	8.6

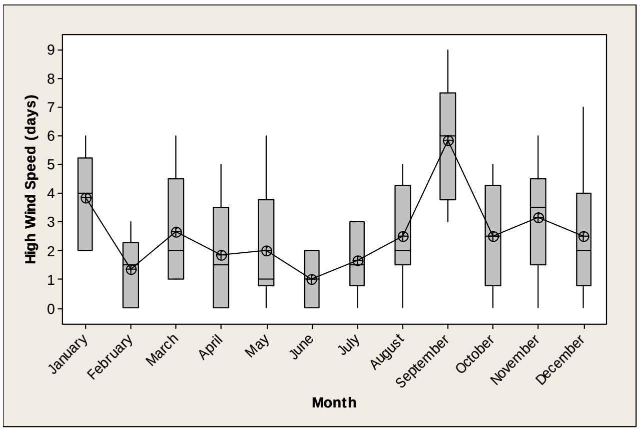

Figure 7. Box plot of the number of days with a high average wind speed (Nov 1999 to Oct 2005)

2.6. Surface Light

At the annual BN review meeting in December 2008, the *Surface Light* node was added to the *Lyngbya* Science network and defined to represent the total available photosynthetically active radiation (PAR) light as measured above the water surface. Populating this node necessitated elicitation of expert opinion to not only define the states and thresholds, but to also propose an alternative measurement to the preferred calculation for surface light as specified by Watkinson *et al.* (2005). The data from their study only spaned a few months and was therefore not suitable for a monthly probability distribution. Consequently we needed to find an alternative measurement which would provide a good approximation for this node. This is a well known and acceptable practice in BN modelling (Borsuk *et al.*, 2006). To this end the cloud cover readings from Beerburrum were used. These readings indicate the level of cloud cover, measured in octets. Therefore the maximum reading for any day with readings at 9am, noon and 3pm is 24. The cloud cover readings were translated into *Adequate* and *Inadequate*, the two expert-defined states of the *Surface Light* node, by comparing the daily average to 14. At the next BN review this approximation of the node and the cut-off can be further debated and changed if required. The resulting CPT table is shown in Table 6 and a box plot of the number of days per month with adequate surface light based on the Beerburrum data from November 1999 to October 2005 is shown in Figure 8.

Table 6. CPT for Surface Light (%)

	Adequate	Inadequate
Sep	88.9	11.1
Oct	67.9	32.1
Nov	67.2	32.8
Dec	59.7	40.3
Jan	64.6	35.4
Feb	56.2	43.8
Mar	63.1	36.9

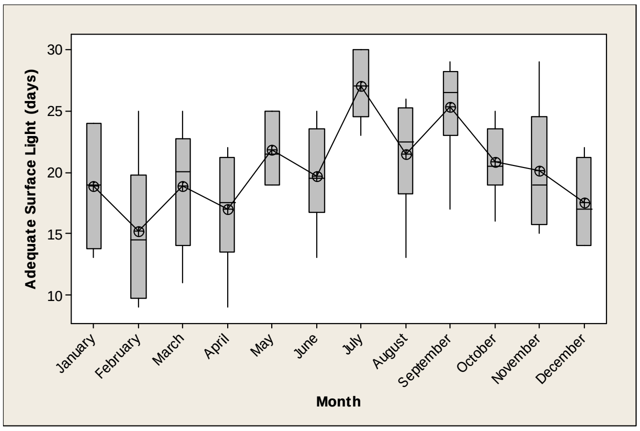

Figure 8. Box plot of the number of days that had adequate surface light, grouped by month (Nov 1999 to Oct 2005)

3. LYNGBYA BLOOM INITIATION

The seven monthly BNs were created using the generic OOBN which represents the reviewed *Lyngbya* Science BN illustrated in Figure 2. The *Lyngbya* Bloom Initiation OOBN for December is shown in Figure 9.

The six input nodes for *Rain - present*, *No of prev dry days*, *Wind Direction*, *Wind Speed*, *Surface Light* and *Temperature* can be seen within the rectangular shaped box with rounded edges. This box is the generic *Lyngbya* Science OOBN. The temporal nodes for December were populated with the CPTs for December (Tables 1 to 6) and then connected to the input nodes of the generic OOBN. Therefore the specific information for December will then flow through the generic *Lyngbya* Science BN taking into account all the interactions modelled in this network to give the probability of a *Lyngbya* bloom initiation for December. From the static *Lyngbya* Science BN the probability of a *Lyngbya* Bloom initiation is 25.3% and for December this increases to 26.8%. The other monthly OOBNs of interest were compiled and run in the Hugin® package and the resulting probabilities are shown in Figure 10. We can see that the probability of *Lyngbya* bloom initiation was less than 25.3% for the months of September to November, but a bloom initiation was more probable for the months of December to March.

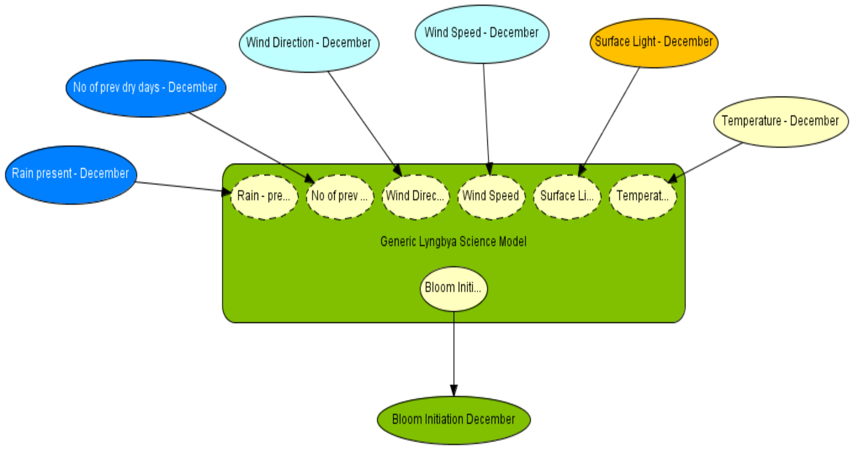

Figure 9. *Lyngbya* Science OOBN for December (Hugin ®)

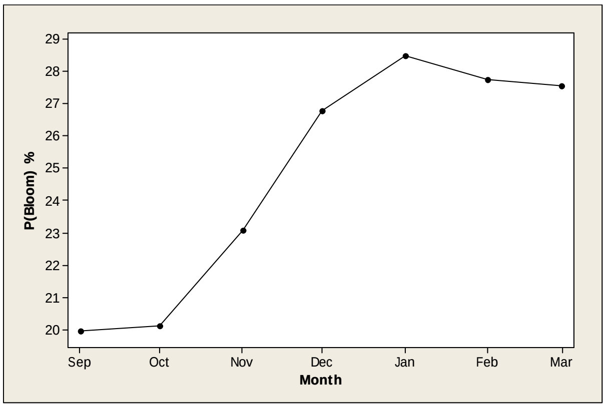

Figure 10. Line graph of the probability of *Lyngbya* bloom initiation from the temporal BNs

This graph clearly demonstrates an increase in *Lyngbya* bloom initiation in the summer months with the rate of increase most notable from October to November and November to December and slowing down but still increasing from December to January. Thereafter the probability, although still elevated, flattens out.

The results presented here are generally in keeping with experts' expectations for *Lyngbya* bloom initiation behaviour; although the elevated probabilities in February and March are perhaps somewhat surprising in that the drop in the bloom initiation probability from January to February and March is not more rapid.

4. DISCUSSION AND CONCLUSIONS

Substantial amounts of expert opinion can be used to inform a static BN. Typically this includes: helping define BN model structure; populating CPTs in the absence of data; and specifying thresholds of nodes populated either by empirical data or expert opinion. We demonstrated here the use of object oriented (OO) Bayesian networks to investigate the temporal behaviour of *Lyngbya* bloom initiation and that expert opinion elicited for a static BN can be reused and extended for a dynamic BN. Two aspects of OO modelling used in this study were reuse and subclassing. We were able to transform the static *Lyngbya* Science BN very simply into an OOBN so that it could be reused for the monthly time slices. Furthermore by using the specific data for each month of interest we 'subclassed' the OOBN to take on the specific values for that month (Koller and Pfeffer, 1997) while retaining the elicited values for all the non-temporal nodes. Expert opinion was necessary to adjust the thresholds for *Rainfall*, *No of prev dry days* and *Temperature* to suit a monthly rather than an annual time scale. Additional information for the *Wind Direction*, *Wind Speed* and *Surface Light* nodes was elicited for the dynamic OOBN. This included thresholds for the nodes and interpreting and summarising the available daily readings (including missing data) to populate the CPTs.

Hodgkinson and Cox (2005) observed that *Lyngbya* blooms, although occurring in both dry and wet months, only occurred in a dry month if it was preceded by a wet period or if there was already a bloom present. They also noted that the majority of blooms occurred after a dry period which was followed by wet periods sometimes with a lag effect. The next stage of modelling *Lyngbya* bloom initiation in a more dynamic way would be to include these lag effects so that the influence of one month flows through to the next. This connectivity between sub-networks allowing information flow between them is an important characteristic of OOBN modelling and one which ought to be elicited from the experts so that this behaviour can be represented in the seven time slices.

Moreover we recommend that the time slices are considered in more detail by the science experts with the purpose of expanding temporal behaviour for a particular month if deemed necessary. In other words, instead of only populating the input nodes of the generic network with data specific to that time slice, additional nodes and interactions could be added which are specific only to that time slice. This is a typical subclassing activity in OOBN modelling (Koller and Pfeffer, 1997).

The inclusion of temporal aspects of *Lyngbya* bloom initiation in this study has provided an improved representation of the risk of algal blooms in the months of interest to stakeholders. However utilising more of the characteristics of OOBN modelling as outlined here would enable *Lyngbya* bloom initiation to be modelled more precisely. Moreover this study illustrates the ease with which expert knowledge can be transferred from a single static model to provide simple dynamic models of ecological processes, such as *Lyngbya* bloom initiation in Moreton Bay. However we identify the need for further elicitation or 'sub-classing' to target the temporal-dependent behaviour of the blooms to provide more precise dynamic models.

ACKNOWLEDGEMENTS

Financial assistance was provided by the Environmental Protection Agency and Australian Government through the South East Queensland Healthy Waterways Partnership, the ARC Centre for Dynamic Systems and Control, and QUT Institute for Sustainable Resources. We fully acknowledge the contributions of the Lyngbya Science Working Group and the BN Review team. We wish to thank three anonymous reviewers for their constructive comments.

REFERENCES

Borsuk, M.E., Reichert, P., Peter, A., Schager, E., and Burkhardt-Holm, P. (2006), Assessing the decline of brown trout (*Salmo trutta*) in Swiss rivers using a Bayesian probability network. *Ecological Modelling*, 192, 224-44.

Hamilton, G., McVinish, R., and Mengersen, K. (2007a), Bayesian model identification and averaging for coastal algal bloom prediction. (unpublished).

Hamilton, G.S., Fielding, F., Chiffings, A.W., Hart, B.T., Johnstone, R.W., and Mengersen, K. (2007b), Investigating the Use of a Bayesian Network to Model the Risk of *Lyngbya majuscula* Bloom Initiation in Deception Bay, Queensland. *Human and Ecological Risk Assessment*, 13, 1271-1279.

Hodgkinson, J. and Cox, M. (2005), Lyngbya blooms in relation to temperature, rainfall, SOI and tides: 2000 to 2005. Moretonbay Waterways and Catchment Partnerships.

Koller, D. and Pfeffer, A. (1997), Object-Oriented Bayesian Networks. Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97), Providence, Rhode Island, 1-3 August 1997, 302-313.

Watkinson, A.J., O'Neil, J.M., and Dennison, W.C. (2005), Ecophysiology of the marine cyanobacterium, Lyngbya majuscula (*Oscillatoriaceae*) in Moreton Bay, Australia. *Harmful Algae*, 4, 697-715.