

Integrating Bayesian Networks and a GIS-based Nutrient Hazard map of *Lyngbya majuscula*

Queensland Coastal Conference 2009

Waves of Change

Tuesday 12 – Friday 15 May 2009
Sea World Resort, Gold Coast

Sandra Johnson¹, Kerrie Mengersen¹, Col Ahern², Shane Pointon², Chris Vowles², Kathleen Ahern³

¹ Queensland University of Technology GPO Box 2434 Brisbane QLD 4001

² Natural Resources and Sciences, NRW, Indooroopilly QLD 4068

³ Environmental Protection Agency PO Box 15155 City East QLD 4002

INTRODUCTION

There is an abundance of expert modelling software available to the environmental scientist and manager. Consequently there are a multitude of models and simulation outputs focussing on a particular aspect of an ecological issue of concern. The integration of the results and the consolidation of knowledge captured in these disparate modelling software systems is a common problem facing the environmental worker.

Here we look at three models of the harmful algal bloom, *Lyngbya majuscula*. Two Bayesian networks were created to model the scientific and management factors in *Lyngbya* bloom initiation and a GIS based model was produced to assess the relative hazard of the nutrients of concern in the catchment, believed to affect the growth, duration and severity of a *Lyngbya* bloom.

We suggest a flexible process to integrate these models and demonstrate this process by an example scenario. The outcome of this scenario is expressed as a probability of *Lyngbya* bloom initiation. We then discuss ways of further automating this process and highlight where the points of integration and flexibility are.

BACKGROUND

Lyngbya majuscula is a cyanobacterium (blue-green algae) occurring naturally in tropical and subtropical coastal areas worldwide (Dennison *et al.*, 1999; Arquitt and Johnstone, 2004). Deception Bay, in Northern Moreton Bay, Queensland, has a history of *Lyngbya* blooms (Watkinson *et al.*, 2005; Ahern *et al.*, 2007), and forms a case study for this investigation.

The South East Queensland (SEQ) Healthy Waterways Partnership, a 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 (Abal *et al.*, 2005). 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 (SEQ Healthy Waterways Partnership, 2007) and as such is more management focussed.

As part of the first phase of the project, Science and Management models for the initiation of a *Lyngbya* bloom was built using Bayesian Networks (BN). The structure of the Science BN was built

by the *Lyngbya* Science Working Group (LSWG) which was drawn from diverse disciplines (Hamilton *et al.*, 2007). The BN was then quantified with annual data and expert knowledge. The Management network was constructed during a series of meetings of the *Lyngbya* Management Working Group (LMWG) with members from local and state government and private organisations. It is a graphical representation of the catchment area showing the nutrients of concern being released into the Bay and identifying the point and diffuse sources discharging these nutrients.

The second phase of the project identified the need to produce a Hazard map of the nutrients of concern Organic Carbon (OC), Phosphorous (P), Bio-available Iron (Fe) and Nitrogen (N)) known to affect the growth, extent and duration of a *Lyngbya* bloom. The soil types and groundwater pH levels are also included, as they affect the solubility of the nutrients and the ability of nutrients to leach from the soil. The Hazard map is a GIS-based model and has a hazard rating for each unique land parcel, obtained by combining the hazard ratings on five GIS layers.

Closer integration of these three models is required to consolidate the available knowledge and to improve management of the catchment area, thereby providing a better understanding of the impact of changes in management practices in the catchment area on *Lyngbya* bloom initiation, growth and duration. By integrating the Hazard map with the BNs previously developed, we can predict the probability of a *Lyngbya* bloom initiation in the Bay for a scenario of interest whilst taking on board the information captured in the Nutrient Hazard map.

METHODS

In the Nutrient Hazard map the hazard ratings for every uniquely identified land parcel in the catchment are calculated by assigning a hazard rating (1 – 4) to each of the nutrients of concern in each of the six GIS coverages. The coverages are acid sulfate soils, landuse, soil type, groundwater, vegetation pre-clearing and vegetation post clearing. The full methodology and rationale for these coverages are given in Pointon *et al.* (2008). These six new layers are then superimposed over each other and values added to produce a merged layer for each nutrient (e.g. merged Fe layer). The new merged layer is then multiplied by a proximity to streams factor for that nutrient to produce a final hazard layer for that nutrient (note: the proximity factor may vary for each nutrient). The final layer for each nutrient is then overlain and added up to produce a final model or hazard model/map (see Fig.1 and 2; Pointon *et al.*, 2008).

In addition, there are two Bayesian networks (BN) for *Lyngbya* bloom initiation prediction. One is the Science BN which focuses on nutrient and physical factors that were agreed by the LSWG to be the most influential contributors to the initiation of *Lyngbya*. The other is the Management BN which focuses on management inputs that potentially influence the delivery of nutrients to the Bay. The 14 land uses of interest identified by the LMWG are six point sources (aquaculture, composting, on-site sewage, poultry, waste disposal and waste water treatment) and eight diffuse sources (agriculture, artificial development, developing and clearing, extractive industries, forestry, grazing, natural vegetation and stormwater).

The challenges to integrate these models are

1. Mapping the land uses in the Hazard maps (74) to the point and diffuse sources in the *Lyngbya* management model (14)
2. Assigning a hazard number to each of the nutrients for each of the point and diffuse sources of the management model, which will be referred to as the 'management hazard factor'
3. Translating the resulting management hazard values into probabilities in the conditional probability tables (CPT) of the Science BN for each of the land uses, to observe the effect on the probability of a *Lyngbya* bloom initiation
4. Translating and quantifying scenarios, e.g. changing from one land use to another, or a percentage change in a particular nutrient of concern

1. Mapping Land Uses to Point and Diffuse sources

On the Hazard Map there are 74 different land uses (LU). In contrast, there are only 14 land uses of interest identified by the LMWG and included in the Management network. The Hazard LUs need to be mapped to the Management LUs to provide communication between the models.

2. Assigning a Management hazard number to each source

There are inherent hazards associated with the various activities in the catchment area and changes to those practices have implications on the possibility of *Lyngbya* blooms initiating in the Bay, as well as the severity, extent and duration of the bloom.

The integration of the models has information flowing from Hazard Map → Management BN → Science BN via the LU mapping. The direction of information flow can also be reversed by applying the LU mapping in reverse. However, since the Hazard to Management LU mapping is 'many to one' it then becomes a 'one to many' mapping and the information has to be split according to some criteria. The most appropriate way would be by apportioning the values using the area of the land parcel for that hazard land use divided by the overall area for the management land use. However the option exists to use other criteria deemed to be suitable.

A Java program has been written to calculate a hazard value for each of the nutrients identified in the Hazard map via the 'Hazard LU to Management LU' map from the previous step. The hazard for each nutrient for a particular land use in the Hazard map is multiplied with the 'coast and stream buffer distance' hazard and the area of each unique land parcel. This value is then divided by the total area. In other words the total hazard is weighted by the relative size of the land parcel with that land use and its coastal proximity. The flexibility exists to change the way in which the combined nutrient hazard value is calculated for each land parcel.

The program then translates the calculated hazards for that hazard land use to a management land use as dictated by the 'Hazard LU to Management LU' map. Since the mapping is many to one, there may be several rows for one management land use. All the rows for each management land use are added together, weighted by area to arrive at the final 'management hazard' for that management land use. These calculated values constitute the set of 'nutrient hazard factors' for that management land use and the resulting table is shown in Table 1 below.

Once again we have an opportunity to change the way in which the program calculates the overall hazard. For example, instead of weighting by area, weighting by the underlying hazard land use may be preferred. In other words even though several hazard land uses map to one management land use, it may be felt that certain hazard land uses contribute more to the overall hazard of that management land use.

Table 1 : Nutrient hazard factors for management land uses

Land Use	Area (km ²)	Fe	P	OC	N	pH
aquaculture	4.97	6.32	5.28	10.04	5.28	1.76
composting	4.17	2.11	1.62	2.61	1.62	1.62
on-site sewage	927.27	2.22	3.10	2.60	3.10	1.55
poultry	7.43	8.89	5.20	9.54	5.20	1.73
waste disposal	7.76	6.47	5.46	10.27	5.46	1.82
waste water treatment plant	156.32	0.59	0.82	1.14	0.82	0.78
agriculture	1,050.31	3.54	4.39	4.63	4.39	2.78
artificial development	513.68	2.91	2.40	3.11	2.40	1.71
developing and clearing	1,201.21	3.25	3.25	3.34	4.73	1.78
extractive industries	89.94	6.93	1.63	2.77	1.63	4.76
forestry	1,641.19	3.98	2.21	5.34	1.11	3.32
grazing	12,571.69	1.11	1.28	1.81	1.28	1.22
natural vegetation	4,096.52	2.42	1.50	4.79	1.41	2.77
stormwater	180.67	2.46	1.43	2.49	1.43	1.60

3. Translating Management Hazard values into probability changes

The *Lyngbya* Science BN reflects the current situation for the nutrients of concern and their contributions to the pool of dissolved nutrients, which is represented by the summary node Available Nutrient Pool (Dissolved). We therefore need to equate the current probabilities of the states of this node, 'Enough' (32.05%) and 'Not enough' (67.95%) to the current overall nutrient hazard as determined by the Hazard Nutrient Map.

The overall hazard for a particular management land use should combine all the nutrients of concern, including pH and soil type. The hazard factors are assumed to be of equal importance and hence a simple average is used to calculate the overall hazard of each management land use. If certain nutrients, pH and/or soil type are shown to have a greater or lesser impact, we can adjust the calculation in the program to be a more complex average which takes into account the relative impact of each hazard factor (nutrients of concern, pH and soil type).

However we first need to derive a single hazard value for the Hazard Map, representing the current situation for nutrients of concern. This value will form the base line for integration with the *Lyngbya* Science BN and is calculated by weighting the hazards according to the area (or another criteria of choice) of each management LU relative to the overall size of the catchment area. The resulting overall hazard value is 2.02. This absolute value does not have any direct meaning as such, but instead should be interpreted in relation to other such hazard factors.

A nutrient hazard rating in the Hazard map may be construed to mean the perceived hazard that there will be 'enough' of that nutrient to cause an increase in growth, extent and duration of a *Lyngbya* bloom. The maximum hazard for any management land use is 16, which results from a nutrient hazard rating of 4 for the land use and 4 for the proximity rating. This represents the absolute certainty of a severe bloom (100% probability). Also the overall management hazard, when viewed relative to the maximum value, needs to be equated to the 32.05% probability of the Available Nutrient Pool (Dissolved) node in the *Lyngbya* Science BN. Then any nutrient hazards and changes in the current situation can be assessed relative to the current situation.

Applying the same logic for each of the management land uses, results in Table 2 below, showing the average hazard rating and the corresponding probabilities for having enough of the dissolved nutrients available to cause a bloom. The latter value can then be used in the Science BN for scenario testing.

Table 2: Management land use showing the average hazard from the hazard map and the corresponding probability of having enough dissolved nutrients in the available pool

Land Use	Average management risk	Probability of 'enough' nutrients
aquaculture	5.74	67.92%
composting	1.92	35.94%
on-site sewage	2.51	43.87%
poultry	6.11	69.77%
waste disposal	5.90	68.73%
waste water treatment plant	0.83	11.62%
agriculture	3.95	57.02%
artificial development	2.51	43.77%
developing and clearing	3.27	51.53%
extractive industries	3.54	53.88%
forestry	3.19	50.85%
grazing	1.34	25.54%
natural vegetation	2.58	44.59%
stormwater	1.88	35.45%

4. Translating and quantifying scenarios

In order to investigate the effect different scenarios have on *Lyngbya* bloom initiation, growth and duration, we need to establish a base line from which these scenarios can be evaluated. As mentioned previously the current probability of having enough dissolved available nutrients in the Science BN is 32.05% and the probability of *Lyngbya* bloom initiation is 25.11%. We will use this bloom initiation probability as the reference point for any scenario testing.

It is possible to adjust the likelihood of a node of interest (such as the bio-available iron nutrient node or the available nutrient pool (dissolved) node) in the Science BN to be the same as the value calculated from the Hazard Map. All the interactions and directions of relationships in the BN are retained and the probabilities are adjusted based on the 'findings' entered for that node. This process of updating the likelihoods of the other nodes is referred to as belief propagation. So as we find out more about a nutrient of concern, we are able to update the BN with this new information and examine how it affects the other key factors in the network and ultimately the effect on the probability of *Lyngbya* bloom initiation.

From Table 2 we can see that if we are interested in a certain nutrient of concern, we can use the corresponding probability in the Science BN to examine how it affects the probability of a *Lyngbya* bloom initiation.

RESULTS

To illustrate how scenario testing may be performed, we use the hypothetical situation of changing the management land use from natural vegetation to agriculture throughout the catchment area. We are interested to see how this change may affect the probability of *Lyngbya* bloom initiation.

To simulate this hypothetical scenario, we need to apply the hazard values for 'nutrients of concern' for agriculture to the areas assigned as natural vegetation. We therefore need to change the hazard values for a total area of 4,097 km². Once this has been done, we recalculate the overall hazard value. The new overall hazard value is 2.27 (previously 2.02) and consequently the probability of 'enough' nutrients of concern for this scenario is 40.83%, a fairly substantial increase compared to the previous value of 32.05%. When we update the BN with this new value and propagate the belief through the network, the probability of a *Lyngbya* bloom initiation increases from 25.11% to 31.34%. The effect of this land use change is diluted by the fact that the proportion of the catchment designated as natural vegetation is only 18.24%.

To better appreciate the potential impact of that type of land use, it is interesting to disregard area and use instead the probabilities from Table 2 for there being 'enough' of the Available Nutrient Pool in the BN for a particular land use. We can then observe the change in probability of *Lyngbya* bloom initiation between agriculture and natural vegetation when this information is propagated through the network. Following this approach causes a change in the probability of a *Lyngbya* bloom initiation from 34.01% for natural vegetation to 42.84% for agriculture.

It is also possible to make some scenario changes in the Hazard map, map the land uses to management land uses and repeat the procedure above. This will then finally give the change in the likelihood of a *Lyngbya* bloom initiation.

DISCUSSION

The integration process is currently partially automated, but it may be desirable to further automate the process if it is being used as a tool to assess impact of changes in the catchment or for business as usual. A graphical user interface (GUI) which enables the dynamic mapping of land uses and then provides output in report or spreadsheet format for the different management land uses would be beneficial. This requires the subsequent automated invocation of the BN java application program interface (API) so that the overall hazard, which have been translated into likelihoods, may be propagated through the network to yield the change in the probability of a

bloom initiation. Furthermore, a GUI which allows scenario entry will provide the end-user with a powerful tool in discussions and negotiations about management actions in the catchment.

CONCLUSIONS

This paper describes and demonstrates the integration of two Bayesian Networks and a GIS-based model for *Lyngbya majuscula* to overcome a common problem faced by environmental researchers in combining outcomes from disparate expert modelling systems. By using a program interface, there are several opportunities to customise the way in which this integration occurs. It also provides the option of automating the interface so that the different ways of integration and scenario testing may be performed by an environmental manager or other person interested in assessing the effects of different management actions on an environmental issue of concern.

TAKE HOME MESSAGE

Bayesian networks with interfacing bespoke Java programs are able to provide a flexible approach to achieve the integration of multiple ecological models constructed for an issue of concern, such as *Lyngbya* bloom initiation. This culminates in a powerful tool to assess the impact of changes to the discharge of nutrients of concern as a result of a change in land use, climate or any other scenario of interest.

ACKNOWLEDGEMENTS

This research was funded by the SEQ Healthy Waterways Partnership and the EPA's Coastal Algal Bloom (CAB) Action Plan. The contributions of the Lyngbya Science and Management teams are fully acknowledged.

REFERENCES

- Abal E G, Greenfield P F, Bunn S E and Tarte D M 2005 Healthy Waterways: Healthy Catchments – An Integrated Research/Management Program to Understand and Reduce Impacts of Sediments and Nutrients on Waterways in Queensland, Australia, *Frontiers of WWW Research and Development - APWeb 2006*, (Harbin, China: Springer Berlin / Heidelberg) pp 1126-35
- Ahern K S, Ahern C R, Savige G M and Udy J W 2007 Mapping the distribution, biomass and tissue nutrient levels of a marine benthic cyanobacteria bloom (*Lyngbya majuscula*) *Marine and Freshwater Research* **58** 883-904
- Arquitt S and Johnstone R 2004 A scoping and consensus building model of a toxic blue-green algae bloom *System Dynamics Review* **20** 179-98
- Dennison W C, O'Neil J M, Duffy E J, Oliver P E and Shaw G R 1999 Blooms of the cyanobacterium *Lyngbya majuscula* in coastal waters of Queensland, Australia *Bulletin de l'Institut Oceanographique, Monaco* **19** 501-6
- Hamilton G S, Fielding F, Chiffings A W, Hart B T, Johnstone R W and Mengersen K 2007 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** 1271-9
- Pointon S M, Ahern K S, Ahern C R, Vowles C M, Eldershaw V J and Preda M 2008 Modelling land based nutrients relating to *Lyngbya majuscula* (Cyanobacteria) growth in Moreton Bay, southeast Queensland, Australia *Nature - Memoirs of the Queensland Museum* **54** 377 - 90
- SEQ Healthy Waterways Partnership 2007 South East Queensland Healthy Waterways Strategy 2007-2012 : Coastal Algal Blooms Action Plan.
- 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

