

**Article Title: Using Bayesian networks to Predict Risk to Estuary Water Quality and Patterns of Benthic Environmental DNA in Queensland**

Authors: Scarlett E. Graham<sup>1,3</sup>, Anthony A. Chariton<sup>2</sup> and Wayne G. Landis<sup>1\*</sup>

\*To whom correspondence may be addressed.

<sup>1</sup> Western Washington University

<sup>2</sup> Macquarie University

<sup>3</sup> Current address: Whatcom Conservation District

\*Corresponding author: Wayne G. Landis

Address: Institute of Environmental Toxicology, MS 9180

Huxley College of the Environment

Western Washington University

Bellingham, WA 98225

Telephone: 360-650-6136

Citation

Graham SE, Chariton AA, Landis WG. In press. Using Bayesian networks to predict risk to estuary water quality and patterns of benthic environmental DNA in Queensland. *Integr Environ Assess Manag*. DOI: 10.1002/ieam.4091

## ABSTRACT

Predictive modeling can inform natural resource management by representing stressor-response pathways in a logical way and quantifying the effects on selected endpoints. This study demonstrates a risk assessment model using the Bayesian network-relative risk model (BN-RRM) approach to predict water quality and; for the first time, eukaryote environmental DNA (eDNA) data as a measure of benthic community structure. Environmental DNA sampling is a technique for biodiversity measurements that involves extracting DNA from environmental samples, amplicon sequencing a targeted gene, in this case the 18s rDNA gene which targets eukaryotes, and matching the sequences to organisms. Using a network of probability distributions, the BN-RRM model predicts risk to water quality objectives and the relative richness of benthic taxa groups in the Noosa, Pine, and Logan estuaries in South East Queensland (SEQ), Australia. The model predicts Dissolved Oxygen more accurately than the Chlorophyll-a water quality endpoint, and photosynthesizing benthos more accurately than heterotrophs. Results of BN-RRM modeling given current inputs indicate that the water quality and benthic assemblages of the Noosa are relatively homogenous across all sub risk regions, and that the Noosa has a 73 – 92 percent probability of achieving water quality objectives, indicating a low relative risk. Conversely, the Middle Logan, Middle Pine, and Lower Pine regions are much less likely to meet objectives (15 – 55 percent probability), indicating a relatively higher risk to water quality in those regions. The benthic community richness patterns associated with risk in the Noosa are high Diatom relative richness and low Green Algae relative richness. The only benthic pattern consistently associated with the relatively higher risk to water quality is high richness of fungi species. The BN-RRM model provides a basis for future predictions and adaptive management at the direction of resource managers.

Key Words: water quality, environmental DNA (eDNA), Bayesian network, ecological risk assessment

**Key Points**

1. We demonstrated that it is possible to use case learning to build a Bayesian network using the Bayesian network relative risk model to predict the probability of meeting water quality objectives for estuarine regions.
2. It is also possible to use the case learning to predict the patterns of occurrence of key eukaryotic taxonomic groups using environmental DNA (eDNA) as the identifier.
3. Once the Bayesian network has been built and the predictions cross validated, it is possible to input the desired endpoint state and then calculate the required environmental conditions.
4. New eDNA data combined with environmental monitoring datasets of sufficient size and quality can be used to predict the probability of the future structure of ecological communities.

**Note to Readers:** There are a number of tables and models called out in the text that can be found in the Supplemental Information. The Models are listed described in the text and represent each of the study areas. 1, Figures as Figure S1, and the Tables as Table S1, etc. References only cited in the Supplemental Information section are listed separately from those in the body of the paper. Examples of the Bayesian network models in Netica<sup>TM</sup> have also been uploaded as part of the Supplemental Information, and notes on how to view them are included in the body of this text.

**Node Terminology:** The names of nodes are capitalized to differentiate them from the organisms or other variables. For example; the node that represents the class of organisms collectively called fungi and represents the diversity of this community is “Fungi”. Similarly, the node that represents the salinity of the surface water is Surface Water Salinity (ppt). The nodes are broken down into two to five discrete states as described in the methods.

## INTRODUCTION

Scientists can contribute to natural resource management by developing predictive models and assessments that link climatic and anthropogenic stressors to environmental and biologic response. Once developed, the models can inform decision-making by estimating risk to valued endpoints and predicting the effects of management actions on valued ecological resources (Barton et al. 2012). A variety of new tools are available to build such models.

Baird et al. (2016) called for the use of the information in large ecological datasets to derive patterns and build causal models to predict environmental effects and manage the environment. Ecological risk assessment, Bayesian networks (BNs) and other tools were recognized by van den Brink et al. (2016) as a means to achieve this goal. Recent advances in sequencing technology and bioinformatics provide an exciting opportunity to collect large datasets of benthic fauna that were previously limited by specialized taxonomy and statistical power (Chariton et al. 2010; Creer et al. 2010; Baird and Hajibabaei 2012). A wide range of benthos are known to respond to environmental gradients and environmental DNA (eDNA) sampling can provide information for further understanding of taxa sensitivities to natural and anthropogenic stressors. For example, in estuaries the photosynthesizing protists are expected to increase as nutrient loads increase (Cloern 2001). Fungi and other organisms associated with the breakdown of organic matter are also expected to increase with nutrient loading and eutrophication, while oxygen-consuming organisms like meiofauna are expected to decrease (Cloern 2001). Environmental DNA sampling can test these hypotheses in a quick and accurate fashion that assesses the entire benthic community, not just those organisms observable by traditional taxonomy (Chariton et al. 2010; Creer et al. 2010; Baird and Hajibabaei 2012). Putting environmental DNA into context

with management goals and other more conventional information about the ecosystem is the goal of this project.

### *Bayesian networks*

Bayesian networks are graphical models that use conditional probabilities to describe relationships between model variables (Marcot 2012; Norsys 2014). They are comprised of nodes and linkages, which represent the variables and cause-effect relationships respectively. Using prior knowledge and data, BNs can calculate the probability of a specific response occurring as well as the associated uncertainty. They are particularly useful for ecosystem scale modeling because they easily integrate many different types of information from different research efforts into a single integrated model (Varis et al. 1994; Borsuk et al. 2004).

Bayesian networks are tools that are often used in ecological modeling and more recently in risk assessment to inform natural resource decision-making (Marcot et al. 2006; McCann et al. 2006; Barton et al. 2012). Havron et al. (2017) demonstrates how BNs can be used in the determination and the portrayal of uncertainty in mapping marine habitat suitability. The visualization of the probability of habitat suitability is presented as a series of maps particularly useful for a resource manager.

### *Ecological Risk Assessment and the Relative Risk Model (RRM)*

Ecological risk assessment provides a useful conceptual framework to organize relationships between environmental variables in context of management goals. In a risk assessment, the management goals are used to define the endpoints and ultimately drive the assessment (Landis

and Wiegiers 2005; Suter 2007) The size of the South East Queensland (SEQ) region warrants a framework that can incorporate multiple stressor interactions across habitats, space and time.

The RRM has been used for 20 years for landscape scale risk assessments to quantify the relative risk for multiple endpoints across sub regions of a site (Wiegiers et al. 1998; Landis and Wiegiers 2005; Ayre and Landis 2012; Hines and Landis 2014; Herring et al. 2015). The RRM framework connects sources, stressors, habitats and impacts and emphasizes the importance of location (Landis and Wiegiers 2005).

Most recently, the RRM has been calculated using BNs because of the flexibility of the modeling approach, the probabilistic nature of the calculations, and the inherent representation of uncertainty (Ayre and Landis 2012). The combination of BNs and the RRM is referred to as a BN-RRM model. The BN-RRM has been used in variety of ecological contexts including to examine the risk of storm water runoff to Coho salmon in the Puget Sound (Hines and Landis 2014), risk to nonindigenous species introduction to Padilla Bay, Washington (Herring et al. 2015), and risk to ecological and human health at a legacy mercury site in the South River, Virginia (Landis et al. 2017, Johns et al. 2017). Landis et al. (2017a, 2017b), and Harris et al. (2017) have demonstrated the usefulness of the BN-RRM in describing the risks to ecosystem services and how these approaches apply to adaptive management.

### *Environmental DNA (eDNA)*

Cordier et al. (2017) demonstrated how eDNA (metabarcoding) and supervised machine learning (SML) can be applied to evaluate the status of marine environments. A key conclusion is that it is possible to use eDNA to evaluate species diversity in marine environments and to assess impacts. They concluded that the combination of eDNA and SML can be equal to or even

surpass current methods that rely on typical taxonomic approaches to determine marine impacts. In our study machine learning and eDNA are integrated into ecological risk assessment using the Bayesian network derivative of the relative risk model.

For the present study, 18S eukaryotic eDNA molecular taxonomy data were used to describe the benthic community endpoints. Researchers at the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia's national science agency, have been collecting and sequencing eDNA from estuarine sediments to assess changes to eukaryotic biotic communities (Chariton et al. 2010; Chariton et al. 2014; Chariton et al. 2015). This *en mass* sampling method allows for identification of thousands of unique sequences per sample, and those sequences can be matched to organisms via online databases. Benthic eukaryote eDNA data collected from SEQ estuarine sediments in 2010 and 2012 were used for this project.

### *Introduction to the South East Queensland Study Sites*

Human populations in coastal areas are changing the water quality and biota of the world's estuaries. In Australia, more than 81% of the 22 million people live within 50 km of the coast (ABS 2003), and the Australian Bureau of Statistics (2003) predicts that the coastal population will increase faster than other regions adding another 1 million people in the next 15 years. Intensive land use for development and associated stressors to water resources are also increasing in coastal areas. Southeast Queensland is Australia's fastest growing region and the combination of heavy rainfall and intensive land uses causes increased loading of non-point source nutrients, organic matter, and suspended sediment to waterways (Ryan et al. 2003; Bunn et al. 2005; Moss et al. 2006). As a result, eutrophication symptoms of depressed oxygen levels and algal blooms are observed in estuaries (Bunn et al. 2005; EHMP 2007). Looking ahead,

models predict that the climate in SEQ is shifting towards higher temperatures and increased frequency of extreme wet and dry events (EHMP 2007; Smith et al. 2013). Climate change may reinforce eutrophication processes by increasing nutrient loading and temperatures (Moss 2011). Thus, given a growing population and a changing climate, using modeling tools like the BN-RRM to predict the changes to estuarine water quality and biota provides valuable information for management of the region (Bunn et al. 2005; Moss et al. 2006).

### *Study Objectives*

The objectives of this research were to:

- 1) Develop an integrated ecological risk assessment model that predicts both water quality risk and benthic taxa (eDNA data) in SEQ estuaries;
- 2) Compare the model predictions for water quality and benthic communities between estuaries and sub regions of the estuaries;
- 3) Evaluate the relevance of incorporating eDNA into a risk assessment framework for the purposes of natural resource management.

We used the BN-RRM approach to quantify the risk of meeting water quality objectives and to predict benthic communities for three estuaries in SEQ; the Noosa, the Pine, and the Logan.

### *Summary of Findings*

We successfully applied the BN-RRM to an integrated assessment of water quality and benthic community structure in three estuaries of SEQ, Australia. Risk to achieving regional water quality objectives was calculated using site-specific monitoring data to quantify relationships between climate, land use, water quality, and benthic communities. To build the

model, we used a large monitoring dataset and machine learning algorithms within Netica (Norsys 2014), the Bayesian network (BN) software, to define conditional probabilities between variables in the model. Monitoring data included water quality, land use, rainfall and benthic environmental DNA (eDNA) data from sediments. The application of eDNA to risk assessment is largely unexplored; and this is the first attempt to synthesize DNA-derived measurements of biological composition into an ecological risk assessment framework.

The three major findings of this work are:

- 1) We demonstrated the use of the BN-RRM approach to develop models that describe the relationships between stressors, water quality, and benthic endpoints in SEQ estuaries. Case learning was used to parameterize relationships in the BN model between land use, water quality, and biota. The structure of the model can be used to test future land use management scenarios and to predict risk to additional endpoints in SEQ estuaries.
- 2) Model results predict that the lower sub regions of the estuaries (nearest the mouth) are more likely to meet water quality objectives than the middle or upper sub regions of the estuaries. The BN-RRM predicted Dissolved oxygen (DO) saturations more accurately than Chlorophyll-*a* (Chl-*a*) concentrations. Photosynthesizing benthic taxa groups, like Diatoms and Green Algae, were predicted more accurately than non-photosynthesizing organisms like Fungi or Meiofauna.
- 3) Environmental DNA data were incorporated into the BN-RRM risk assessment framework as the relative richness of six benthic taxa groups. This approach models the patterns of benthic fauna response to water quality stressors. Future work to determine management goals for benthic fauna and to incorporate other measures of community assemblage and function would enhance this assessment.

The BN-RRM models created for this study provide a basis for managers to understand current conditions, predict future states of endpoints, and ultimately use for decision-making.

## METHODS

This section first provides a summary of the BN-RRM methodologies and then explains how they were applied to this risk assessment for SEQ estuaries. A detailed description of eDNA sampling, sequencing, and analysis of the eDNA data is provided the Supplemental Information Section 2.

### *Description of the South East Queensland (SEQ) study sites*

The SEQ region centers around the Queensland state capital of Brisbane and Moreton Bay (Figure 1). The Noosa catchment forms the northern border, and the Queensland-New South Wales state border is to the south (Bunn et al. 2005). The waterways of the SEQ include 14 major river catchments, which flow from west to east discharging into either Moreton Bay or the Pacific Ocean.

Figure-1. Overview map of the Southeast Queensland region, study sites (Noosa, Pine and Logan catchments), and land use classifications.

Moreton Bay is a large shallow embayment, separated from the ocean by sand islands, and it accumulates sediment, nutrients, and pollutants from the catchments that drain into it (Bunn et al. 2005; SEQHWP 2007a). The bay is a designated marine park and includes a wetland of international significance under the Ramsar Convention for protection of wetland habitats and migratory birds (Abal et al. 2005). Compared with other large embayments around the world,

Moreton Bay is relatively quickly flushed (50-55 day residence time in the central bay), though in the southern and western zones, flushing times are longer (66-75 day residence time for the Logan River) (Dennison and Abal 1999; SEQHWP 2007a).

The BN-RRM model in this study focuses on two classes of stressors to SEQ estuaries: climate and land use. These are two of the largest stressors influencing the environmental condition in SEQ estuaries. In terms of management changes, land use is the major factor that can be managed.

*Climate.* The SEQ climate is subtropical with mild dry winters (June – August) and hot, humid, and rainy summers (December – February). Rainfall varies widely between the seasons and from year to year, with rainfall during dry years less than half that of wet years (SEQHWP 2007a). In the summer and autumn months, heavy rainfalls result in high seasonal flows often with flooding in SEQ waterways. Future climate change projections predict that rainfall variability is likely to increase yet total rainfall is likely to decrease by 10 to 30 percent (SEQHWP 2007a).

Record rainfall fell from December 2010 through January 2011 during a strong La Niña cycle, causing the second highest flooding in Brisbane and surrounding areas since the beginning of the 20<sup>th</sup> century (van den Honert and McAneney 2011). Rainfall in the 600 to 1,200 mm range was widespread along most of the Queensland coast (van den Honert and McAneney 2011). One of the impacts of this flooding event was a temporary increase in suspended sediment from erosion, which resulted in the extensive deposition of fine mud in Moreton Bay and the surrounding estuaries (O'Brien et al. 2012).

*Land use.* The SEQ region supports a rapidly growing population of 2.7 million people who use the waterways for drinking water supply, commercial and recreational fisheries, and other

recreational activities (SEQHWP 2007a). Human activity since European settlement has significantly changed the landscape with only one quarter of the remnant vegetation remaining intact (Bunn et al. 2005).

According to the Australian Land Use and Management Classification scheme, the SEQ region is predominately classified as production from relatively natural environments, which includes grazing (56%) and conservation land (16%) secondary land use types. Areas of production from relatively natural environments predominates in upper to mid catchment areas. In the lower catchment nearest the estuaries and coast, intensive uses (shown in red in Figure 1) predominate. Intensive land uses are known to highly inhibit natural processes, and are associated with complete or nearly complete removal of remnant vegetation (Queensland Government 2010). Examples include intensive horticulture, animal husbandry, industrial, residential and farm infrastructure, utilities, mining, and waste treatment and disposal (State of Queensland 2014). Since 1999, the intensive uses have increased by 9% as more land is being developed to meet the needs of a growing population (State of Queensland 2014).

Run-off events from intensive land use areas tend to be larger in volume and carry with them larger amounts of non-point source pollutants including sediment, nitrogen and phosphorus (Cottingham et al. 2010). The estimated loadings of pollutants per unit area of urban land use significantly higher than from rural sources (twice as much for sediment and up to seven times as much for nitrogen) (Abal et al. 2005). Previous studies have used percent intensive land use as predictors of sediment and nutrient loading to waterways in a risk assessment context (Moss et al. 2006; Scheltinga and Moss 2007). Thus, we are also using percent intensive use in our model.

### *Management of SEQ waterways*

Management of the SEQ waterways is a joint effort between the Queensland Government, local governments, and the South-East Queensland Healthy Waterways Partnership (SEQHWP), a non-profit organization formed in the 1990s between government, industry, universities, and community stakeholders. The Healthy Waterways Partnership focuses on the conservation of the region's water resources and is concerned with future water security (in terms of both quality and quantity) in the face of a changing climate and a growing population (Bunn et al. 2005). The partnership manages the Ecosystem Health Monitoring Program (EHMP), a comprehensive ambient monitoring program, on behalf of its member organizations. Through the EHMP, water quality and biologic monitoring data have been collected routinely from freshwater, estuarine and marine sites for the past 15 years. Using an index calculated with the EHMP data, the Healthy Waterways Partnership produces an annual report card for the freshwater and estuarine portions of each of the 14 SEQ catchments. The report card grades (A through F) are meant to communicate the current condition of each catchment and can be compared from year to year (Bunn et al. 2005).

The three SEQ estuaries evaluated in this risk assessment, the Noosa, Pine and Logan (Figure 1), represent a range of land use impact, water quality, and morphology in the SEQ. We considered only the estuarine portion of each waterway, which extends from the marine limit (either Moreton Bay or the Pacific Ocean in SEQ) to the tidal limit in the rivers. Estuary boundaries, as well as the lower, middle and upper reaches with the estuaries, were delineated by the State of Queensland and made available via GIS shapefiles. This research relies heavily on EHMP monitoring and spatial data provided by the State of Queensland. A brief summary of the

environmental conditions of each estuary is provided in Table 1 and a description of each estuary follows.

Table 1. Environmental Description of each Estuary.

*Noosa.* The Noosa catchment is located on the northern border of the SEQ region with headwaters and much of the upper catchment in the Great Sandy National Park (Figure S1). The catchment forms a coastal lagoon system of five lakes and discharges directly into the Pacific Ocean at Noosa Heads, which is a popular beach for surfing, fishing and tourism (Sunshine Coast Council and Queensland Government 2012). There are no point sources discharging directly to estuarine waters. The Healthy Waterways Partnership regularly gives the Noosa highest grades (Table 1) and considers the overall environmental condition of the Noosa excellent (SEQHWP 2015).

*Pine.* While the Pine catchment is similar in overall area to the Noosa, the estuarine section is smaller than the Noosa (Table 1, Figure 1). The Pine includes two tributaries, the North and South Pine Rivers (Figure S2), which are located to the north of the city of Brisbane. Both rivers originate from undeveloped regions of protected forest land and flow east through rural residential areas before entering a highly urbanized area near the estuary mouth and discharging into Moreton Bay (SEQHWP 2015). The North Pine River has been dammed, forming Lake Samsonvale, which provides drinking water and is also a recreational resource (Pine Rivers Catchment Association Inc. 2002). The Murrumba Downs Sewage Treatment Plant (STP) discharges into the North Pine River approximately 10 km from the estuary mouth and has not been upgraded since 2000 when EHMP monitoring began. The estuary flows into Hayes Inlet

that contains mangrove and seagrass habitat that is important to recreational and commercial fisheries, as well as migrating birds. The Healthy Waterways Partnership gave the Pine a C grade in 2015, which is consistent with previous years, and considers the environmental conditions to be fair (SEQHWP 2015).

*Logan.* The Logan is located just south of the city of Brisbane is the second largest catchment in the SEQ (Figure S3). The mouth of the Logan estuary lies at the southern end of the Moreton Bay, and is not as well flushed as the Noosa or Pine estuaries, which impacts nutrient levels and oxygen renewal in estuaries. The catchment supports a diverse array of land use including agriculture, grazing and dairying in the upper catchment and residential urban areas in the lower catchment along the estuary. In addition, several aquaculture facilities are located along the banks of the Lower Logan risk region near the mouth, and discharge periodically into the waterway (SEQHWP 2015). The Middle Logan risk region receives treated sewage and waste water directly from the Loganholme Wastewater Treatment Plant (last upgraded in 2014; located ~17km from estuary mouth), and indirectly from the Beenleigh Water Reclamation Facility (~14km from estuary mouth), which discharges into the Albert River before it confluences with Logan. The Healthy Waterways Partnership gave the Logan a D grade in 2015, although historically scores an F, and considers the environmental conditions to be poor (SEQHWP 2015).

### **Formulation of the BN-Relative Risk Model**

The RRM methodology described in Landis and Wiegers (2005) was used for selecting risk assessment endpoints and developing the conceptual model. Additional methods outlined in Ayre and Landis (2012) describe the use of BNs within the RRM framework for modeling and risk calculations.

The first step in the RRM is to solicit ecological values from stakeholders and collaboratively decide on endpoints for the risk assessment. The selected endpoints should have management goals associated with them so the risk assessor can develop the model as specifically as possible. Next, all possible stressors to the endpoints and the sources of those stressors are identified through further solicitation and research (Landis and Wieggers 2005). Ultimately, only the most important stressors and sources relating to the endpoint are retained for the risk assessment. The RRM uses a multiple stressor approach that considers both anthropogenic and natural/environmental stressors (Landis and Wieggers 2005). Location is very important to the assessment; stressors, habitats, and endpoints are mapped as information is gathered about the site(s). Spatial analysis is also used to delineate the site into sub risk regions based on similar environmental attributes or environmental resource management objectives (Landis and Wieggers 2005).

A conceptual model linking sources of stressors to endpoints is then constructed based on causal relationships between the variables in the system (Landis and Wieggers 2005). All variables are defined with an appropriate measurement attribute (% DO saturation or mm of rainfall per 30-day period). The variables are then discretized into ranked states and risk is calculated as the probability distribution of achieving a given state. Evaluation of the model includes a sensitivity analysis and comparison to observed conditions if available. Finally, model results are communicated in a fashion that portrays the relative risk to endpoints in the context of management goals (Landis and Wieggers 2005).

### *Selection of Endpoints*

Ideally, risk assessors and environmental modelers host multiple stakeholder meetings to identify management goals, select valued ecological endpoints, and receive feedback while developing the conceptual model (Borsuk et al. 2004; Landis and Wieggers 2005). Given the currently available information, limitations in funding, time, and the location of the study sites; we relied on a less formal process that included a literature review and one-on-one meetings with CSIRO, SEQ Healthy Waterways Partnership, and Queensland government scientists. Stakeholder values had been previously solicited through other research efforts, and the results of those surveys were publicly available (Abal et al. 2005; SEQHWP 2007a; SEQHWP 2007b; Healthy Waters, EHP 2013). Based on that information, we chose to focus on water quality and benthic biota as endpoints for this risk assessment. Given the flexibility of BN modeling, future SEQ risk assessments can easily incorporate additional endpoints such as macro fauna, human health, or ecosystem services into the models that we have developed for this study.

Water resources are very important to SEQ, and the Healthy Waterways Partnership has hosted workshops with stakeholders to classify the ecological value of waterways (SEQHWP 2007a; Healthy Waters, EHP 2013). The Noosa estuary, as a tourist destination that abuts a national park, is designated as High Ecological Value by stakeholders, while the urbanized Pine and Logan estuaries are designated as Slightly to Moderately Disturbed. Highly Disturbed is the lowest possible value designation, though no estuaries in the SEQ are designated as such. Estuaries with higher ecological value have different water quality objectives than those with a lower designation, and the relative risk calculations will reflect these differences.

In a risk assessment framework, endpoints are defined as entities and attributes, where attributes describe the valued qualities of the entity (USEPA 1998). In SEQ, the regional Water

Quality Objectives are the entity and DO and Chl-*a* are the attributes. For the biotic endpoint, the entity is the benthic community assemblage as measured by the eukaryotic eDNA, and the attribute is the relative richness of a suite taxa groups including Diatoms, Dinoflagellates, Fungi, Meiofauna, Protozoans, and Green Algae.

Dissolved oxygen and Chl-*a* were selected as endpoints for this risk assessment because they are the two most common water quality responses to increased nutrient loading, are symptoms of eutrophication, and can negatively affect other estuarine biota and human health (Cloern 2001). Increases in phytoplankton growth, as measured by Chl-*a*, is a primary symptom of eutrophication and increases in Chl-*a* cascade into secondary symptoms like DO depletion from microbial degradation. When phytoplankton blooms are too dense, they cause a range of effects from further reducing DO levels to producing toxins to decreasing water clarity. Water that appears murky or is closed for swimming or recreating due to toxic algal blooms has negative effects on recreation and public perception (ANZECC/ARMCANZ 2000). Aquatic animals, including fish and benthic invertebrates, require oxygen to breathe, and depleted DO levels reduce abundance by forcing organisms to relocate, causing direct mortality, or toxic inhibition of submerged aquatic vegetation (Cloern 2001; Nezlin et al. 2009).

The benthic eukaryote community, as identified by eDNA sequencing techniques, is the biotic endpoint in this risk assessment. Benthic communities are commonly monitored in waterways, and many studies demonstrate their responsiveness to changes in environmental conditions (Johnston and Roberts 2009). Benthos including eukaryotes (nematodes, protists, fungi, etc.) and prokaryotes (bacteria) inherently underpin all trophic levels, and their biodiversity as well as functional attributes can inform managers about the state of an ecosystem (Kennedy and Jacoby 1999). There are currently no regional objectives to assess risk of

achieving for these endpoints, so instead the results reflect the patterns of benthic assemblages as a result of stressors to each estuary. The eDNA eukaryotic benthic dataset includes over 8,000 unique molecular operational taxonomic units (MOTU) sequence reads. All of the taxa associated with the MOTU cannot be represented in a BN efficiently and effectively with one node for each taxa. To overcome this challenge, the richness of the most common and frequent taxa groups were chosen to represent the benthic community assemblage.

### *Identification of Stressors and Sources of Stressors*

The SEQ Healthy Waterways Partnership has identified sediment and nutrient loading and reduced environmental (natural) freshwater flows as the major impacts affecting the SEQ waterways (SEQHWP 2007a; SEQHWP 2007a; State of Queensland 2009). To a lesser extent, toxicants including pesticides and heavy metals have also been identified as a source of anthropogenic stress to the region's waterways (Bunn et al. 2005). However, after review of 2012 surface sediment data collected by CSIRO, pesticides and heavy metals were not detected or detected at concentrations below management goals, so they were not included as major stressors to water quality and biota in this assessment. The low contaminant loadings were likely due to the extensive flushing caused by an extensive flood event in early 2011. Excess sediment and nutrient in waterways were retained as stressors for the BN-RRM.

Currently, the major source of nutrient and sediment loading is diffuse runoff from intensive land uses, which include agriculture and exposed hill slopes used for grazing, horticulture, and intensive animal production (SEQHWP 2007a; SEQHWP 2007a; State of Queensland 2009; Cottingham et al. 2010). Point sources including poorly functioning sewage treatment plants and aquaculture discharges have been managed in the last 15 years, and contribute nutrient loading to

a lesser extent than non-point source land uses (Cottingham et al. 2010). The major natural sources contributing nutrients and suspended sediments in estuarine waters are climate (rainfall and temperature) and ocean water influence (measured by salinity). These three environmental sources (rainfall, salinity, and season) plus the anthropogenic source (intensive land use) have been included as inputs to the BN-RRM model. The inputs for each risk region were based on site-specific data.

### *Risk Region Delineation*

Only the estuarine reaches of each of the three catchments are considered in this study, and the estuarine boundaries were delineated based on tidal limits by the State of Queensland. We divided the estuaries into sub risk regions to capture the differences within each catchment. This regionalization process took many attributes of the estuaries into account including (in order of importance):

- management goals (i.e. water quality objectives for Lower Noosa estuary versus the Middle Noosa estuary based SEQ documentation);
- relationships between variables in the estuary;
- salinity gradients;
- land use; and
- location of point sources.

For example, in the Noosa and Pine estuaries, the State of Queensland sets different water quality objectives for the lower (more saline) versus middle (fresher water) sections of the estuaries, so those regions were delineated based on management goals and salinity gradients. The Logan estuary, on the other hand, only has one set of water quality objectives for the entire estuary so other factors were used for regionalization. Figures S1-S3 present the different risk regions of the estuaries, 2012 land use designations, point sources, and EHMP and eDNA sample locations. The eDNA and EHMP locations are co-located in most instances.

### *Model Structure*

The structure of the BN-RRM model for this study (Figure 2) was developed based on the original RRM framework (Landis and Wieggers 1997, 2005) with sources of stressors linking to habitats, and habitats linking to effects on endpoints. The pathways flow from left to right and variables in the BN-RRM are called nodes (Figure 2). The linkages between the nodes indicate a causal or correlative relationship between variables (Marcot 2012). These relationships were determined using a variety of information including a literature review of estuarine water quality science, regional reports for South East Queensland estuaries, EHMP data, sediment eDNA data, and feedback from CSIRO scientists.

Figure 2. Bayesian network relative risk conceptual model method model and Bayesian network for the Logan estuary with the Middle Logan risk region stressors selected.

Stressors and sources of those stressors are the input nodes and start the pathway on the left side of the model (Figure 2). The input nodes predict the intermediate water quality nodes (Total Nitrogen [TN], Total Phosphorous [TP], Turbidity, and Temperature). The intermediate nodes link to the endpoint nodes, which describe the predicted impact to the attributes of the valued ecological entity.

A BN-RRM model strives to balance accuracy, parsimony and relevance to management and risk assessment, and therefore does not include all possible variables that affect the endpoints. The BN-RRM model constructed for this risk assessment includes the most important measurable variables that predict water quality endpoints (DO and Chl-*a*) and benthic community endpoint in a single integrated model. The model was built in Netica (Norsys 2014).

A total of seven BN models were constructed, one for each risk region: Lower Noosa, Middle Noosa, Upper Noosa, Lower Pine, Middle Pine, Lower Logan and Middle Logan (Figure 2).

### *Model Assumptions*

All estuary models for this study have exactly the same physical BN structure, meaning that we assume the flow of cause-effect relationships between variables in the SEQ estuaries are the same. This works because the models were built based on the breadth of information about chemical and biological relationships in estuaries. Further, the models must have the same structure (and discretization of variables) to make the model results comparable between sub risk regions and estuaries.

### *Data, Discretization and Construction of the Conditional Probability Tables*

The BN-RRM method uses many types of data to discretize variables, parameterize the conditional probability tables (CPTs), and define the inputs for each risk region. Data used in the model, as related to these three categories, are described below.

#### Discretizing the Nodes.

Each node in the model was discretized into discrete states (Table 2). The goal of discretizing was to represent the influences of the variables on the endpoint with the fewest discrete states necessary. In most cases, we created four states corresponding to the zero, low, med, high risk ranking scheme described by Landis and Wieggers (2005), and used by many other BN-RRMs (Ayre and Landis 2012; Herring et al. 2015; Landis et al. 2016 Jan). For some nodes, three or five states were preferable to four states to more accurately reflect the data and to compare between regions. Where possible, discrete states were established using classifications

recognized by estuary science or local SEQ management objectives and justifications are provided in Table 2. For water quality variables that have a local management objective (TN, TP, Turbidity, Chl-a, DO), the lowest possible risk states achieves the objective, and all other states exceed the objective. Dissolved oxygen is slightly different because it is possible to exceed and fall below the objective percent saturation (Table 2).

**Table 2.** Methodology used to discretize model variables, and the states of these variables

The discrete states of all input, intermediate, and benthic community nodes are exactly the same for all models and estuaries. However, the two water quality endpoint nodes (DO and Chl-*a*) vary slightly between regions based on differences in regional management objectives. We used a wide variety of information sources to discretize the nodes, including peer-reviewed literature, governmental reports, water quality objectives, natural breaks in the data, or a combination of each (Table 2).

*Derivation of the conditional probability table (CPT).*

The relationship between two or more parent nodes connected to a child node is defined by a CPT. Conditional probability tables are represented as a matrix of probabilities of a child node state occurring state given the state of its input nodes (Norsys 2014). Conditional probability tables can be parameterized using a variety of methods. These methods can be broken down into four categories: expert judgment, empirical evidence, mathematical or biological equations, and case file learning (Marcot et al. 2006; Pollino et al. 2007; Chen and Pollino 2012). In a single

model, CPTs for different nodes may be completed using different methods (Chen and Pollino 2012) or combination of methods may be used within a single CPT (Pollino et al. 2007).

The CPTs in the SEQ BN-RRMs were parameterized using one method: case files and the expectation-maximization (EM-learning) algorithm, an automated function available in Netica (Norsys 2014). We chose this method based on Lucena-Moya et al. (2015), who also used case learning to parameterize predictor nodes for benthic endpoints. A case file is a compilation of a set of data that go together to provide information about the variables (nodes) in the model. For example, water quality data combined with weather data collected on the same day. The EM-learning algorithm iteratively calculates the maximum likelihood estimates for the node states in the model given the case file data and the model structure. Expectation-maximum learning has been routinely used for other environmental BN modelling (Pollino et al. 2007; Lucena-Moya et al. 2015), and was selected over other algorithms because it deals well with missing data.

We used a unique case file for each estuary model and to parameterize the CPTs. The three case files (one for each estuary) consisted of SEQ EHMP monitoring data that had been collected on a monthly basis from 1999-2014 at regular locations within the estuary (Figures S1-S3). Each record (or case) in the case file was matched to the corresponding previous 30-day rainfall total and % Intensive land use ( $n = 5,032$ ,  $n = 6,204$ ,  $n = 3,621$  cases for the Noosa, Logan, and Pine respectively). These case files were used to parameterize the CPTs between all nodes except for the benthic richness endpoints.

The benthic endpoint nodes of the model were parameterized using a single separate benthic data case file for all estuaries ( $n = 287$ ). A single file was used because the eDNA dataset for individual estuaries was relatively small. Using benthic samples from different estuaries to

parameterize CPTs is consistent with Lucena-Moya et al. (2015) who also used BNs and case-learning to predict ecological assemblages.

*Setting the estuary specific input (stressor) distributions.*

The input distributions for the Salinity and Intensive Landuse nodes were set using risk-region specific data based on the sub region of estuary (lower, mid or upper). Salinity was determined based on EHMP monitoring data collected from 1999-2014, and the Intensive Land use was calculated from the 2012-2013 Queensland land use data. Rainfall input data were only specified as an estuary specific, not down to the sub-region and based on observations by the Queensland Bureau of Meteorology from 1985-2015 for the nearest weather station to the estuary.

*Benthic eDNA data*

Scientists from CSIRO, Australia's National Science Agency, collected the sediment samples and sequenced the eDNA data used in this risk assessment. The surface sediment eDNA samples were collected by CSIRO during two sampling events (Summer 2010 and Fall 2012) from five estuaries in SEQ: the Noosa, Maroochydore, Pine, Logan and Currumbin. While only three estuaries (the Noosa, Pine, and Logan) were evaluated in the risk assessment presented in this manuscript, eDNA from all five sampled estuaries were used for determining the water quality predictors in the model and for parameterizing the CPTs. A summary CSIRO's methods and our analyses used to determine predictors is provided in the Supplemental Information section 2 Benthic environmental sampling and analysis.

### *Model Evaluation*

We evaluated the BN-RRM models using three methods. These methods were used both during the model development process and to evaluate the final models. The results of these model tests can be used to further guide BN development (Marcot et al. 2006; Marcot 2012).

### *Predicted Versus Observed*

A simple first test is to determine whether trends in model predictions are consistent with field observations (Pollino et al. 2007). One of the outputs of a BN model is the mean or expected value for a given node. For each risk region, we compared the BN mean value to the observed average value from the EHMP or eDNA data for the water quality and benthic endpoints respectively.

### *Cross Validation*

Bayesian network model cross validation was performed for each estuary and endpoint with the Netica feature *Test with Cases* (Norsys 2014). The purpose of this feature is to grade a BN using a set of real cases to see how well the predictions match the actual. We used 80% of the case file data for a given estuary to build or train the model and the remaining 20% to test the model (Pollino et al. 2007; Chen and Pollino 2012). The error rates were compared across endpoints and between estuaries and are presented in the results section under cross validation. An example of a cross validation file (PINE\_LOWER\_CROSSVALIDATION.neta) can be found in the supplemental information for this publication.

### *Sensitivity analysis*

Sensitivity analysis quantifies how much the distribution of an endpoint node is influenced by the probability distributions of the other nodes (Pollino et al. 2007; Marcot 2012). This analysis can be used during model development to identify errors in CPTs or model structure

(Pollino et al. 2007). The variables that the model is most sensitive to should be supported by the primary literature about the system or by empirical evidence. Once the model is completed, sensitivity results can provide guidance for future data collection by identifying which variables are most prediction of changes to the endpoint. A sensitivity analysis was performed for each endpoint in each risk region using the Sensitivity to Findings feature in Netica (Norsys 2014). Because the variables are discretized into states, sensitivity was measured as mutual information, or reduction in variable entropy (Marcot 2012).

## RESULTS

### *Bayesian Network Relative Risk Method (BN-RRM) models*

The BN-RRM model for each estuary has two layers of predictions (Figure 2). The first layer uses climate stressors (Season, Monthly Rainfall, and Salinity) and the Intensive Land use stressor to predict four intermediate water quality effects (TN, TP, Turbidity, and Temperature). The second layer of the model uses the four water quality effects to predict the response of the eight risk assessment endpoints (DO, Chl-*a*, Diatom, Dinoflagellates, Green Algae, Fungi, Meiofauna, and Protozoans). The intermediate water quality variables are the effects of the stressors on the surface water habitat, and this pathway reflects the RRM framework.

The BN model results for the endpoints are summarized in three ways. First there are the predicted probabilities of the endpoint states occurring (i.e. 74% likelihood that the Lower Noosa will meet water quality objectives for DO). These probabilities sum to 100% and are called the posterior probability distributions (PPDs). The PPD communicates the most likely state for the endpoint and the uncertainty for that prediction. Uncertainty, in this context, pertains to the dispersion of the probability values among the endpoint states, that is, the spread of the possible predictions (Marcot et al. 2006). In Netica, the PPDs are shown as horizontal bar charts, and we

have summarized the PPDs for all endpoint nodes (Figures 3, S4). A second model result is the probability that the water quality objective is achieved in a given risk region or estuary (Table 3). A high probability of achieving the objectives ( $\geq 75\%$ ) is associated with low relative risk and conversely low probabilities ( $< 50\%$ ) are associated with high relative risk. Third, there is an expected value associated with each node, which is the predicted mean value weighted by the probability of occurrence (Norsys 2014). For the endpoint nodes, the expected values are either the mean concentration of the water quality endpoint or the mean richness for the benthic endpoint. The mean values are useful for comparing between risk regions. To check the accuracy of the model, we have compared the expected value to the observed mean value for each risk region as part of the model evaluation process. The following sections describe BN-RRM results for the water quality and benthic endpoints, as well as model evaluation results including sensitivity analysis and cross validation.

**Table 3.** BN model results for the probability of achieving regional water quality objectives for each risk region.

#### *Water Quality Endpoints – Posterior Probability Distributions*

For both the DO and Chl-*a* endpoints, the % probability of achieving water quality objectives is higher in the Noosa estuary than in the Pine and Logan estuaries (Table 3). Within the Noosa estuary, the Lower, Middle and Upper sub regions have very similar PPDs and high certainty of achieving the objective risk states) (between 73-91% probability of achieving the Objective risk states for both DO and Chl-*a*) (Table 3). In the Logan and Pine estuaries, however, the sub regions have different PPDs and there is greater uncertainty of achieving a particular state (i.e.

the spread of the PPD is larger). In general, the Lower Logan and Lower Pine sub regions are more likely to achieve Objectives for DO and Chl-*a* than the Middle sub regions (Table 3). In the Middle Logan and Middle Pine regions, the Chl-*a* distributions are skewed toward higher Chl-*a* concentrations, while the DO distributions are skewed toward lower DO % sat. (Figure 3).

Figure 3- Examples of the output distributions from the BN-RRM calculation: Water quality and Primary Producers. The state on the Y-axis labeled “Objective” indicates achieving the regional water quality objective. The objectives and states differ slightly between regions. Also shown are the Primary producer (mostly photosynthetic) posterior probability distributions for each region and estuary. No objectives had been set at the time of this study.

In the Lower Logan risk region, the probability of achieving water quality objectives for DO was surprisingly high (69% probability), given that the Logan estuary is regularly given the lowest grades (given a D in 2015) by the SEQ Healthy Waterways Partnership report cards. The Lower Logan has a very wide central basin at the mouth of the estuary, that can contribute to greater renewal of DO in the surface waters via wind and tidal mixing (Ryan et al. 2003; Nezlin et al. 2009).

#### *Benthic Endpoints – Posterior Probability Distributions*

For this report, we have organized the benthic eukaryotic taxa groups into two classes based on their trophic level and function:

- Primary Producers (photosynthetic): Diatoms, Dinoflagellates and Green Algae.
- Primary Consumers and Decomposers (mostly non-photosynthetic): Protozoans, Meiofauna, and Fungi.

Comparisons of benthic richness can be made 1) between estuaries (e.g. which estuary has the highest overall Diatom richness?) and 2) between risk regions (e.g. does Diatom richness change

between the Lower and Middle sub regions of an estuary?). The patterns of richness between estuaries reflect differences in stressors in the estuaries. The Lower risk regions represent marine regions and the Middle/Upper region represent less saline regions.

#### *Comparison between estuaries: Primary Producers*

The primary producer taxa groups have different PPDs between estuaries, with the Diatoms and Green Algae having the largest differences (Figure 3). In the Noosa estuary, Diatom relative richness is likely to be high (in the 15-30% relative richness state), and Green Algae and Dinoflagellate richness are likely to be low (in the 0-2.5 % and 2.5-5% states, respectively). Diatoms were identified as an indicator species for the Noosa estuary by Chariton et al. (2015) based on Threshold Indicator Taxa Analysis (TITAN). Thus, the BN modeling results are consistent with the indicator analysis results.

The Pine and Logan estuaries are more similar to each other. Diatom richness is lower, and Green Algae and Dinoflagellate richness is higher compared to the Noosa (Figure 3). These differences are most prominent in the Middle sub regions of the Pine and Logan. The PPDs for the Pine and Logan are spread across multiple states reflecting a high uncertainty of a particular state occurring.

#### *Comparison between sub regions: Primary Producers*

The Noosa benthic taxa PPDs are similar for all sub regions, meaning there are not many differences along the salinity gradient from the Lower to the Upper risk regions (Figure 3). This same trend was observed in the water quality, with similar high probability of meeting objectives in all sub regions. The Logan and Pine PPDs, however, are different between risk regions, with

Diatom richness decreasing from marine to fresh waters, while Green Algae and Dinoflagellate increases from marine to fresh waters.

#### *Comparison between estuaries: Primary Consumers and Decomposers*

For the non-photosynthesing taxa groups, only the Fungi taxa have markedly different patterns between estuaries. In the Noosa estuary, Fungi relative richness is predicted to be low (in the 0-2.5% richness state) and that prediction is fairly certain ( $\geq 60\%$  probability of the low state occurring) (Figure S4). In the Pine and Logan estuaries, the PPDs show overall higher Fungi richness compared to the Noosa, but PPDs are spread across the multiple risk states reflecting high uncertainty of a given state occurring (Figure S4).

The Meiofauna and Protozoan richness patterns are similar between all regions and all estuaries. In addition, the probability of any one state occurring is often less than 50% (Figure S4), reflecting uncertainty about which richness state may occur and also reflecting the high variability of Meiofauna and Protozoan richness between samples in the estuaries.

#### *Comparisons between sub regions: Primary Consumers and Decomposers*

Only the Fungi taxa richness reveals patterns between sub regions. Along the salinity gradient, Fungi richness increases from marine (Lower sub regions) to fresher waters (Middle/Upper sub regions). This pattern is very evident in the Logan and Pine estuaries, with the Middle Logan distributions shifting to a moderate richness state (2.5-5% richness) being the most likely to occur. In the Noosa, the pattern of Fungi richness increasing from marine to freshwater is also present but to much a lesser extent. Overall, the Noosa PPDs are rather

homogenous between risk regions. This result is similar to patterns observed for the water quality endpoints and photosynthetic benthic taxa groups, and is a major finding in this work.

## **Model Evaluation**

### *Predicted versus Observed*

Observed water quality data for each risk region was plotted against expected mean predictions from the BN-RRM model (Figure 4). Comparisons indicate that the trends between monitoring data and predictions are generally consistent. However, the model predicts slightly higher Chl-*a* concentrations than what is observed (Figure 4). The higher predictions are likely a result of the discretized states for Chl-*a* and the method by which Netica calculates the expected mean value (weighting it by the probability of occurrence). The highest state, ranging from 20 – 65 µg/L (Table 2), is likely skewing the expected mean high.

Figure 4. Model results for water quality endpoints compared to actual observed average concentrations for each risk region. Abbreviations for the regions are shown in Table 3. The shading of the bars corresponds to the three estuaries: white = Noosa, gray = Pine, and black = Logan.

Observed benthic data for each risk region were plotted against the expected mean richness predictions from the BN-RRM (Figure S5). The BN model predictions match the actual observed richness trends more accurately for the photosynthetic groups (Diatoms, Dinoflagellates, and Green Algae) than the non-photosynthetic groups (Protozoan, Meiofauna, and Fungi; Figure S5). Of the non-photosynthetic groups, the BN model predicts Fungi richness trends the best.

### *Cross Validation*

Cross validation error rates for the water quality endpoints ranged from 33 to 50% (Table S1), and error rates were lowest in the Logan estuary models and for the DO endpoint. The Chl-*a* error rates are high for the Pine and Logan and may be because we chose a higher resolution for the lower concentrations in the discretization of the Chl-*a* variable (see Table 2) and the model is not able to predict the difference between 0-2  $\mu\text{g/L}$  and 2-4  $\mu\text{g/L}$  for the more eutrophic estuaries. The Noosa has consistently very low Chl-*a* concentrations, so the model predicts Chl-*a* better in the Noosa and that is reflected in the lower error rates. Error rates for the benthic endpoints were comparable to the water quality endpoints and ranged between 28 and 40%. All of the estuaries have the same error rates for the benthic endpoint because the benthic CPTs were parameterized with the same case file.

The high average error rates across all models and endpoints (>28%) likely reflects the large amount of variability within both the water quality and benthic eDNA data. It is likely that the sequencing of deceased organisms, an artefact of eDNA sampling, was a major contributor to the variability in the benthic eDNA data (Chariton et al., 2015). These error rates are consistent with other BN models that predicted benthic taxa using case-learning to parameterize the CPTs (Lucena-Moya et al. 2015).

### *Sensitivity Analysis*

Sensitivity analysis is used to identify which variables are most influential with respect to endpoint variables. Nodes that are physically closer in the network to the endpoint node will have a greater influence on the endpoints (higher % mutual information) based purely on the structure of the model. Sensitivity analysis was run multiple times throughout the model

development process to evaluate both the structure and discretization of variables in the model. These results are based on the final model configuration. we have divided the sensitivity results into two categories: sensitivity to inputs/stressors (Figure S6), and sensitivity to intermediate water quality variables (Figure S7).

*Sensitivity to Inputs/Stressors.* The Salinity stressor was clearly the most influential input for all endpoints except for Chl-*a*, where Season was more influential (Figure S6). Sensitivities to the Land Use and Rainfall stressors varied between the endpoints, but in general had lower mutual info compared to Salinity (Figure S7). Outcomes of the sensitivity analysis concur with estuarine science that the horizontal salinity gradient and estuarine circulation drives nutrients and DO levels as well as biotic assemblages.

*Sensitivity to Intermediate Variables.* The most influential intermediate water quality variables for all endpoints were the nutrients TN and TP, as these two variables are linked to each endpoint. Total nitrogen had slightly higher mutual information than TP for all endpoints (Figure S7). This is consistent with the finding that nitrogen is the limiting nutrient to primary production in SEQ estuary and marine waters where light is not limiting, (Cottingham et al. 2010).

### **Interactive Capability of Netica Models as Research Product**

This modeling effort has produced seven highly interactive BN models that can be provided to resource managers to understand and communicate the response of endpoints to stressors present in the regions. Netica files (.neta) are also available for download. The ability to test stressor scenarios and predict quantitative changes in the endpoint states can be informative to decision makers. For example, using Netica you can select % Intensive Landuse to be in the 65

to 80 % state for the Middle Logan risk region, and all of the other nodes will automatically update their probability distributions based on that input. The result of higher % Intensive Land use is lower DO, higher Chl-a, lower Diatoms, and higher Fungi.

There are limitations to which stressor scenarios you can test. Currently, it is not possible to make predictions with stressor states where there was no observed data in the case-files. For example, none of the risk regions currently have the highest state of Intensive Land Use (80 to 100%), so the model cannot learn how that state would affect intermediate water quality parameters like TN or TP. With the case-learning method used for parameterizing the CPTs, the results of unknown states are given an even distribution, with all resultant states equally likely to occur. The unknown parts of the CPT can be parameterized in future iterations of these BN-RRM models with other sources of information including primary literature, other model predictions, or expert elicitation.

## DISCUSSION

The objectives of this study were to demonstrate a methodology for integrating eDNA into an ecological risk assessment framework and to then evaluate the results and the usefulness of the method. The SEQ estuary BN-RRM models relatively are simple, but they accurately predict the current patterns of water quality and relative richness of a number of benthic taxa groups. Looking ahead, the variables and relationships in the model can be added to or updated as more information becomes available or specific management questions are asked. Further, while these BN-RRM provide a demonstration of integrating eDNA results into a risk assessment, the

richness metric has a few drawbacks which are discussed below. Future models should try to improve upon these ones.

### *Water quality and benthic endpoint results*

Risk of not achieving water quality objectives is higher in the Pine and Logan estuaries compared to the Noosa, with the highest risk of not achieving the objectives occurring in the middle estuary sub regions. The Noosa has low risk of not achieving objectives in all sub regions, and the endpoint expected values and PPDs do not differ between sub regions. The homogeneity of the water quality of the Noosa is also reflected in the eDNA benthic richness patterns. The Pine and Logan however, have different water quality risk and different benthic richness patterns between the sub regions. The differences between sub regions of the Logan and Pine are due to two variables; 1) salinity and 2) nitrogen levels. Sensitivity analysis identified these two variables as being important in the BN-RRMs.

There are consistent trends in benthic taxa richness for regions that have greater than a 50% chance of achieving water quality objectives for at least one variable (Table 3). These regions include the Noosa and Lower Pine and Logan regions. Trends among the risk regions include:

- Diatoms have high richness and
- Green Algae has low richness.

Thus, these biotic signals in these models predict a high probability of achieving water quality objectives, and subsequently low risk to not achieving objectives.

The Fungi richness patterns also correlate to water quality, but tend to reflect the intermediate water quality variables (nutrients and turbidity) more than the DO and Chl-*a* endpoints. For example, the Noosa estuary, which has the lowest nutrient and turbidity levels, has the lowest Fungi richness. Similarly, the Pine and Logan estuaries, which have much higher

nutrients and turbidity, have higher Fungi relative richness. Fungi taxa are major decomposers of woody and herbaceous substrates in marine and estuarine ecosystems (McGuire and Treseder, 2010). Higher levels of Intense land use in the Pine and Logan estuary catchments are the cause of the higher nutrients, and consequently more organic matter entering the waterways. Thus, Fungi richness is higher in the Pine and Logan because they have a larger supply of material to decompose as a food source.

Relative richness, used as the attribute of the benthic taxa groups in this assessment, has some drawbacks; and BN models and risk assessments that incorporate traditional or DNA-derived biotic data can use other attributes that may better reflect the structure of the community. The first drawback to a richness metric is that there are no established management goals in SEQ for estuarine benthic communities. Management goals are policy choices and should be developed by resource managers and stakeholders. Risk assessors can play a role in the process by developing quantitative integrated risk assessments like the one presented in this study to provide analysis and information to formulate those goals. The second drawback is that a richness metric reduces the community assemblage response into a single value. Some studies have incorporated multiple community metrics into one model (e.g. richness and evenness, diversity, filterer and grazer abundance) (Allan et al. 2012; Leigh et al. 2012). Still, modeling the response of the entire assemblage would provide a more comprehensive understanding of estuarine condition (Lucena-Moya et al. 2015). Moving forward, we propose using environmental distance measurements of communities, similar to those utilized in non-indigenous species risk assessments (Bradie et al. 2015). In the present study, an example would be to evaluate communities based on their distance measure to the Noosa estuary which gets A grades in the EHMP program.

### *Uncertainty and data needs*

BNs explicitly express uncertainty by representing all variables as distributions. The BN-RRM model development process can reveal the current state of knowledge of the system, and call attention to data gaps that should be filled to make management decisions.

One of the largest sources of uncertainty in the model is the eDNA richness. This uncertainty is due to three factors: 1) data were collected from only two sampling events; 2) there is high natural variability of benthic eukaryote communities; and 3) the amplification of deceased organisms, residual fragments of organisms, spores, eggs etc. The latter may be attenuated by examining RNA instead of DNA (Laroche et al. 2017), however, this would greatly increase the costs of obtaining the data. The eDNA data used in this study were collected from two events over the course of three years (February/Summer 2010 and May/Fall 2012) with no replication of seasons. Given that these organisms are ephemeral and sensitive to changes in water quality and labile inputs, more years and seasons of data would further inform the model. Furthermore, because all eDNA data was combined into a single case file to construct the CPTs for the model, model validation results for the eDNA richness were not estuary specific.

On the other hand, there is no paucity of sampling data for water quality or rainfall in SEQ (> 10 years worth), and the relationships between these variables (rainfall, temperature, salinity, TN, TP, turbidity, DO, and Chl-*a*) are likely the strongest in the model. The most likely state of these variables is still uncertain for given inputs, especially in regions other than the lower estuary. This uncertainty is due primarily to the natural variability in water quality itself, not due to a lack of data. The endpoints, DO and Chl-*a*, are known to be highly variable in estuaries and dependent on seasonal variation and even the time of day sampling occurred. Given that the

models were parameterized with over 3,000 cases per estuary, we are confident that the model appropriately captures the variability of intermediate water quality nodes as well as DO and Chl-a endpoint responses.

There is less information available for Intensive Land Use because that measure has been sampled only once every 5-6 years in Queensland. Refining the relationship between land use and water column nutrients (TN and TP) and turbidity would be very useful to managers because land use is a variable that can be managed via policy and engineering measures. Once these relationships are further refined, managers could easily quantify how much management would have to occur to achieve desirable endpoint results.

#### *Integrating environmental DNA into a risk assessment*

While there is still much work to be done to measure the benthic community assemblages, reduce uncertainty, and further refine relationships in the BN-RRM, this study demonstrates that integration of eDNA into risk assessment framework is possible. By using case-learning, the relationships between the eDNA and predictor variables can be quickly determined.

Bayesian network modeling is meant to be an iterative process where information can be added as it becomes known or new data is collected. As new EHMP data is collected each year, more evidence can inform the model. In addition, new endpoints can be assessed using the model because BNs are easily updated with new information. Further, BN models can be used interactively to demonstrate the quantitative changes that cascade from stressors to endpoints.

#### ACKNOWLEDGEMENTS

The environmental data used for this project were accessed from internet based portals administered by the State of Queensland and the South East Queensland Healthy Waterways

Partnership. Dr. Anthony Chariton formally of CSIRO Oceans and Atmosphere in Sydney, Australia supplied the 18S environmental DNA data of benthic eukaryote organisms, as well as provided guidance on data analysis. Meagan Harris provided guidance on the BN-RRM process, as well as a thoughtful review of this document.

Funding for travel to Australia to meet with stakeholders and visit the estuary sites was provided by CSIRO, the Institute of Environmental Toxicology at Western Washington University, Huxley Small Grant, and the Dean's Fund for Sustainability Studies. We have no conflict of interest to declare.

Data availability. The data used in the model are accessible via the CSIRO data portal <http://dx.doi.org/10.4225/08/52DF4D8008B99>. The sources of information used to develop the case files are listed in Supplemental section 3. The .neta models are available as part of the supplemental information included in this manuscript and can be read with the free version of Netica™ available at <https://www.norsys.com/download.html>.

## REFERENCES

- Abal E, Bunn S, Dennison W, editors. 2005. Healthy Waterways, Healthy Catchments: Making the connection in South East Queensland, Australia. Brisbane, Australia: Moreton Bay Waterways and Catchments Partnership.
- ABS. 2003. Regional Population Growth, Australia and New Zealand, 2001-02. [ABS] Australian Bureau of Statistics, Canberra Report No.: ACT cat. no. 3218.0.
- Allan JD, Yuan LL, Black P, Stockton T, Davies PE, Magierowski RH, Read SM. 2012. Investigating the relationships between environmental stressors and stream condition using Bayesian belief networks. *Freshw. Biol.* 57:58–73.
- ANZECC/ARMCANZ. 2000. Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management of Australia and New Zealand. Canberra, ACT (National water quality management strategy; no.4)).
- Ayre KK, Landis WG. 2012. A Bayesian Approach to Landscape Ecological Risk Assessment Applied to the Upper Grande Ronde Watershed, Oregon. *Hum. Ecol. Risk Assess. Int. J.* 18:946–970.
- Baird DJ, Hajibabaei M. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing: NEWS AND VIEWS: OPINION. *Mol. Ecol.* 21:2039–2044.
- Baird DJ, Van den Brink PJ, Chariton AA, Dafforn KA, Johnston EL. 2016. New diagnostics for multiply stressed marine and freshwater ecosystems: integrating models, ecoinformatics and big data. *Mar. Freshw. Res.*, 67, 391–392
- Barton DN, Kuikka S, Varis O, Uusitalo L, Henriksen HJ, Borsuk M, de la Hera A, Farmani R, Johnson S, Linnell JD. 2012. Bayesian networks in environmental and resource management. *Integr. Environ. Assess. Manag.* 8:418–429.

- Borsuk ME, Stow CA, Reckhow KH. 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecol. Model.* 173:219–239.
- Bradie J, Pietrobon A, Leung B. 2015. Beyond species-specific assessments: an analysis and validation of environmental distance metrics for non-indigenous species risk assessment. *Biol. Invasions* 17:3455–3465.
- Bunn S, Abal E, Greenfield P, Tarte D. 2005. Making the connection between healthy waterways and healthy catchments: South East Queensland, Australia.
- Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM. 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Front. Ecol. Environ.* 8:233–238.
- Chariton AA, Ho KT, Proestou D, Bik H, Simpson SL, Portis LM, Cantwell MG, Baguley JG, Burgess RM, Pelletier MM, Peron M. 2014. A molecular-based approach for examining responses of eukaryotes in microcosms to contaminant-spiked estuarine sediments: DNA comparison of assemblages exposed to spiked sediments. *Environ. Toxicol. Chem.* 33:359–369.
- Chariton AA, Stephenson S, Morgan MJ, Steven ADL, Colloff MJ, Court LN, Hardy CM. 2015. Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environ. Pollut.* 203:165–174.
- Chariton AA, Sun M, Gibson J, Webb JJ, Leung KMY, Hickey CW, Hose GC. 2016. Emergent technologies and analytical approaches for understanding the effects of multiple stressors in aquatic environments. *Mar. Freshw. Res.* 67, 414–428
- Chen SH, Pollino CA. 2012. Good practice in Bayesian network modelling. *Environ. Model. Softw.* 37:134–145.
- Cloern JE. 2001. Review: Our evolving conceptual model of the coastal eutrophication problem. *Mar. Ecol. Prog. Ser.*:223–253.

- Cordier T, Esling P, Lejzerowicz F, Visco J, Ouadahi A, Martins C, Cedhagen T, Pawlowski J. 2017. Predicting the Ecological Quality Status of Marine Environments from eDNA Metabarcoding Data Using Supervised Machine Learning. *Environ. Sci. Technol.* 51, 9118–9126
- Cottingham R, Franz Delfau K, Garde P. 2010. Managing diffuse water pollution in South East Queensland: An analysis of the role of the Healthy Waterways Partnership. Brisbane, Queensland: International Water Centre.
- Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML, Lamshead PJD, Thomas WK. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol. Ecol.* 19:4–20.
- Dennison W, Abal E. 1999. Moreton Bay study: A scientific basis for the healthy waterways campaign. South East Queensland Regional Water Quality Management Strategy, Brisbane.
- EHMP. 2007. Ecosystem Health Monitoring Program 2005-06 Annual Technical Report. Brisbane, Australia: South East Queensland Healthy Waterways Partnership (SEQHWP).
- Havron A, Goldfinger C, Henkel S, Marcot BG, Romsos C, Gilbane L. 2017. Mapping marine habitat suitability and uncertainty of Bayesian networks: a case study using Pacific benthic macrofauna. *Ecosphere* .8:1-25.
- Healthy Waters, EHP. 2013. Establishing draft environmental values, management goals and water quality objectives: Under the Environmental Protection (Water) Policy 2009.
- Herring CE, Stinson J, Landis WG. 2015. Evaluating nonindigenous species management in a Bayesian networks derived relative risk framework for Padilla Bay, WA, USA: Risk Assessment for Nonindigenous Species. *Integr. Environ. Assess. Manag.* 11:640–652.
- Hines EE, Landis WG. 2014. Regional risk assessment of the Puyallup River Watershed and the evaluation of low impact development in meeting management goals: Risk Assessment and Low Impact Development. *Integr. Environ. Assess. Manag.* 10:269–278.

- Johns, AF, Graham SE, Harris MJ, Markiewicz AJ, Stinson JM, Landis WG, 2017. Using the Bayesian Network Relative Risk Model Risk Assessment Process to Evaluate Management Alternatives for the South River and Upper Shenandoah River, Virginia. *Integr Environ Assess Manag.* 13:100-114
- Johnston EL, Roberts DA. 2009. Contaminants reduce the richness and evenness of marine communities: A review and meta-analysis. *Environ. Pollut.* 157:1745–1752.
- Kennedy AD, Jacoby CA. 1999. Biological indicators of marine environmental health: Meiofauna - A neglected benthic component? *Environ. Monit. Assess.* 54:47–69.
- Landis WG, Ayre KK, Johns AF, Summers HM, Stinson J, Harris MJ, Herring CE, Markiewicz AJ. 2017. The Multiple Stressor Ecological Risk Assessment for the Mercury Contaminated South River and Upper Shenandoah River Using the Bayesian Network-Relative Risk Model. *Integr Environ Assess Manag.* 13:85-99
- Landis WG, Wiegers J. 2005. Chapter 2: Introduction to the regional risk assessment using the relative risk model. In: *Regional Scale Ecological Risk Assessment Using the Relative Risk Model*. CRC Press Boca Raton. p. 11–36.
- Laroche O, Wood SA, Tremblay LA, Lear G, Ellis JI, Pochon X. 2017. Metabarcoding monitoring analysis: the pros and cons of using co-extracted environmental DNA and RNA data to assess offshore oil production impacts on benthic communities. *PeerJ* 5: e3347
- Leigh C, Stewart-Koster B, Sheldon F, Burford MA. 2012. Understanding multiple ecological responses to anthropogenic disturbance: rivers and potential flow regime change. *Ecol. Appl.* 22:250–263.
- Lucena-Moya P, Brawata R, Kath J, Harrison E, ElSawah S, Dyer F. 2015. Discretization of continuous predictor variables in Bayesian networks: An ecological threshold approach. *Environ. Model. Softw.* 66:36–45.

- Marcot B, Steventon J, Sutherland G, McCann R. 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Can. J. For. Res.* 36:3063–3074.
- Marcot BG. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecol. Model.* 230:50–62
- McCann R, Marcot B, Ellis R. 2006. Bayesian belief networks: applications in ecology and natural resource management. *Can. J. For. Res.* 36:3053–3062.
- McGuire KL and Treseder KK. 2010. Microbial communities and their relevance for ecosystem models: Decomposition as a case study. *Soil Biol Biochem* 42: 529-535.
- Moss A, Cox M, Scheltinga D, Rissik D. 2006. Integrated estuary assessment framework. Indooroopilly, Qld.: Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management (Coastal CRC).
- Moss B. 2011. Allied attack: climate change and eutrophication. *Inland Waters* 1:101–105.
- Nezlin NP, Kamer K, Hyde J, Stein ED. 2009. Dissolved oxygen dynamics in a eutrophic estuary, Upper Newport Bay, California. *Estuar. Coast. Shelf Sci.* 82:139–151.
- Norsys. 2014. NeticaTM. Vancouver, BC, Canada: Norsys Software Corporation.
- O'Brien K, Tuazon D, Grinham A, Callaghan D. 2012. Impact of mud deposited by 2011 flood on marine and estuarine habitats in Moreton Bay. Healthy Waterways.
- Pine Rivers Catchment Association Inc. 2002. Pine Rivers Catchment Management Strategy.
- Pollino CA, Woodberry O, Nicholson A, Korb K, Hart BT. 2007. Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environ. Model. Softw.* 22:1140–1152.
- Queensland Government. 2010. The Australian Land Use and Management (ALUM) Classification Version 7.

- Ryan D, Heap A, Radke L, Heggie D. 2003. Conceptual models of Australia's estuaries and coastal waterways: applications for coastal resource management. Canberra: Geoscience Australia, Dept. of Industry, Tourism & Resources Report No.: GeoCat No. 47336.
- Scheltinga DM, Moss A. 2007. A framework for assessing the health of coastal waters: a trial of the national set of estuarine, coastal and marine indicators in Queensland (CONDENSED VERSION). Environmental Protection Agency Queensland, prepared for the National Land & Water Resources Audit, Canberra.
- SEQHWP. 2007a. South East Queensland Healthy Waterways Strategy 2007-2012. Management Strategy Evaluation Action Plan.
- SEQHWP. 2007b. South East Queensland Healthy Waterways Strategy 2007-2012: Strategy Overview. SEQHWP and the Australian Government Report : ISBN 978-0-646-48515-7.
- SEQHWP. 2015. Healthy Waterways 2015 Report Card. *Healthy Waterw.*  
<http://healthywaterways.org/reportcard/#/zone/89/2015/condition>
- Smith I, Syktus J, McAlpine C, Wong K. 2013. Squeezing information from regional climate change projections - results from a synthesis of CMIP5 results for south-east Queensland, Australia. *Clim. Change* 121:609–619.
- State of Queensland. 2009. South East Queensland Natural Resource Management Plan 2009-2031. Department of Environmental and Resource Management: State of Queensland Report No.: ISBN 9311662183491 #28951.
- State of Queensland. 2014. Land use summary: South East Queensland NRM Region. Brisbane, Queensland: (Department of Science, Information Technology, Innovation and the Arts).
- Sunshine Coast Council, Queensland Government. 2012. Noosa River Catchment and Estuary Management Plan (2012-15).
- Suter GW, editor. 2007. Ecological risk assessment. 2nd ed. Boca Raton: CRC Press/Taylor & Francis.

van den Brink P, Choung CB, Landis W, Pinto MM, Pettigrove V, Scanes P, Smith R, Stauber J. 2016. New approaches to the ecological risk assessment of multiple stressors. *Mar. Freshw. Res.* 67:393–41

van den Honert RC, McAneney J. 2011. The 2011 Brisbane Floods: Causes, Impacts and Implications. *Water* 3:1149–1173.

Varis O, Kuikka S, Taskinen A. 1994. Modeling for water quality decisions: uncertainty and subjectivity in information, in objectives, and in model structure. *Ecol. Model.* 74:91–101.

## Tables

**Table 1. Environmental Description of each Estuary.**

<b>Estuary</b>	<b>Estuary Catchment Area (in km<sup>2</sup>)</b>	<b>Estuary Type<sup>1</sup></b>	<b>Major Energy Influence<sup>1</sup></b>	<b>2015 EHMP Grade<sup>2</sup></b>	<b>Land use and Sources of Anthropogenic Stressors<sup>2</sup></b>
Noosa	250	Wave-dominated Estuary	Wave Energy	A-	Includes mostly natural areas including a National Park. Urban areas occupy < 3% of catchment and are concentrated around lower estuarine reaches. No point sources.
Pine	67	Tide-dominated Estuary	Tide Energy	C	Estuarine reaches area highly urbanized; 2 impoundments forming lakes on North Pine River. Point sources include a WWTP.
Logan	300	Tide-dominated Delta	Tide Energy	D	Estuarine reaches flow through rural residential, urban and industrial areas. Point sources include aquaculture facilities and two WWTP.

## Notes

1. Estuary Type determined by Geoscience Australia (Ryan et al. 2003).
2. Catchment area and land use information summarized from [www.health-e-waterways.org](http://www.health-e-waterways.org).

**Table 2.** Methodology used to discretize model variables, and the states of these variables

<b>Variable</b>	<b>Discretization Methodology and Justification</b>	<b>States</b>
<b>Stressor Nodes</b>		
Land Use (% Intensive Uses)	States were determined by Moss et al. (2006) to predict sediment and nutrients in SEQ waterways based on the intensive land use designation. The Moss et al. (2006) work was done as part of another SEQ estuary assessment framework.	< 30 %
		30 - 50 %
		50 - 65 %
		65 - 80 %
		≥ 80 %
Total Monthly Rainfall	States were determined using natural breaks (Jenks optimization) of 30-day rainfall totals from 1985-2015, and rounded to the nearest 10. Totals were calculated from Queensland BOM stations in each estuary: Logan WWTP (station #40854), Noosa Tewatin (#40908), and Pine Petrie Mill (#40171).	0 - 50 mm
		50 - 100 mm
		100 - 200 mm
		200 - 600 mm
Season	Austral Seasons	Autumn (Sept-Nov)
		Winter (Dec-Feb)
		Spring (Mar-May)
		Summer (June-Aug)
Salinity	Venice System for Classification of Marine Waters (Venice system, 1958).	0 - 0.5 ppt
		0.5 - 5 ppt
		5 - 18 ppt
		18 - 30 ppt
		30 - 37 ppt

<b>Intermediate Nodes</b>		
Total Nitrogen Concentration	The lowest state ( 0.3 mg/L) meets default trigger values for the SEQ region (ANZECC/ARMCANZ 2000). The highest state (1-3.5 mg/L) is classified as a high in a survey in Bricker et al. (2003). The middle states were discretized based on natural breaks (Jenks optimization) of EHMP monitoring data.	<p>&lt; 0.3 mg/L (Objective)</p> <p>0.3 - 0.6 mg/L</p> <p>0.6 - 1 mg/L</p> <p>1 - 3.5 mg/L</p>
Total Phosphorous Concentration	The lowest state (0-0.03 mg/L) meets default trigger values for the SEQ region (ANZECC/ARMCANZ 2000). The next state (0.03-0.1 mg/L) is classified as high in a survey in Bricker et al. 2003, and was set taking into account site-specific EHMP data (TP is higher in SEQ estuaries).	<p>&lt; 0.03 mg/L (Objective)</p> <p>0.03 - 0.1 mg/L</p> <p>0.1 - 0.5 mg/L</p> <p>0.5 - 1.6 mg/L</p>
Water Temperature	The states were discretized based on natural breaks (Jenks optimization) of the EHMP monitoring data.	<p>13 - 19 °C</p> <p>19 - 23 °C</p> <p>23 - 26 °C</p> <p>26 - 31 °C</p>
Turbidity	The lowest state (< 8 NTU) meets default trigger values for SEQ (ANZECC/ARMCANZ 2000). The highest state (>100 NTU) is known to limit phytoplankton growth, and cause fish avoidance (Bisson and Bilby 1982). The (40 - 100 NTU) state was based on Moss et al. 2006.	<p>&lt; 8 NTU (Objective)</p> <p>8 - 40 NTU</p> <p>40 - 100 NTU</p> <p>100 - 1000 NTU</p>
<b>Endpoint Nodes - Water Quality</b>		
Dissolved Oxygen Saturation	The Lower Pine and Lower Noosa Objective states were set to 90-105% sat. and the rest were set to 85-105% sat. based on SEQ regional water quality objectives (State of Queensland 2010). The lowest state (<50% sat.) is	< 50 % sat.

	considered hypoxic (Breitburg 2002). The remaining states were set based on natural breaks (Jenks optimization) that were rounded.	50 - 70 % sat. 70 - 85 % sat. 85 - 105 % sat. (Objective) 105 - 167 % sat.
Chlorophyll-a Concentration	Based on region-specific objectives, the Objective states varies between risk region in the Noosa and Pine estuaries. The states >4 µg/L were set based on low, medium, high and hypereutrophic ranges reported in a survey in Bricker et al. 2003.	< 2 µg/L (Objective)  2 - 4 µg/L (Objective) 4 - 10 µg/L 10 - 20 µg/L  20 - 65 µg/L
<b>Endpoint Nodes - Benthic Relative Richness</b>		
Diatom, Meiofauna, and Protozoan	Benthic endpoints states were discretized based on P. Lucena-Moya et al. (2015) to maximize the detection of community change in response. The Diatom, Dinoflagellate, Meiofauna, and Protozoan have the same four states.	0 - 5 %  5 - 10 % 10 - 15 % 15 - 30 %
Dinoflagellate, Green Algae and Fungi	Dinoflagellate, Fungi and Green Algae were detected at lower relative richnesses, so two very low states (0 - 2.5 % and 2.5% - 5%) were added to replace the highest state (15 - 30%) to more accurately represent the biota richness.	0 - 2.5 %  2.5 - 5 % 5 - 10 % 10 - 15 %

**Table 3.** BN model results for the probability of achieving regional water quality objectives for each risk region.

Risk Regions	Dissolved Oxygen		Chlorophyll-a	
	Objective (% saturation)	% Probability to achieve objective	Objective ( $\mu\text{g/L}$ )	% Probability to achieve objective
Noosa Lower (NL)	90 - 105%	74%	< 1.8 $\mu\text{g/L}$	73%
Noosa Middle(NM)	85 - 105%	82%	< 2.2 $\mu\text{g/L}$	77%
Noosa Upper (NU)	85 - 105%	75%	< 5 $\mu\text{g/L}$	92%
Pine Lower (PL)	90 - 105%	55%	< 2 $\mu\text{g/L}$	26%
Pine Middle (PM)	85 - 105%	26%	< 4 $\mu\text{g/L}$	40%
Logan Lower (LL)	85 - 105%	69%	< 4 $\mu\text{g/L}$	61%
Logan Middle (LM)	85 - 105%	15%	< 4 $\mu\text{g/L}$	49%

Figures

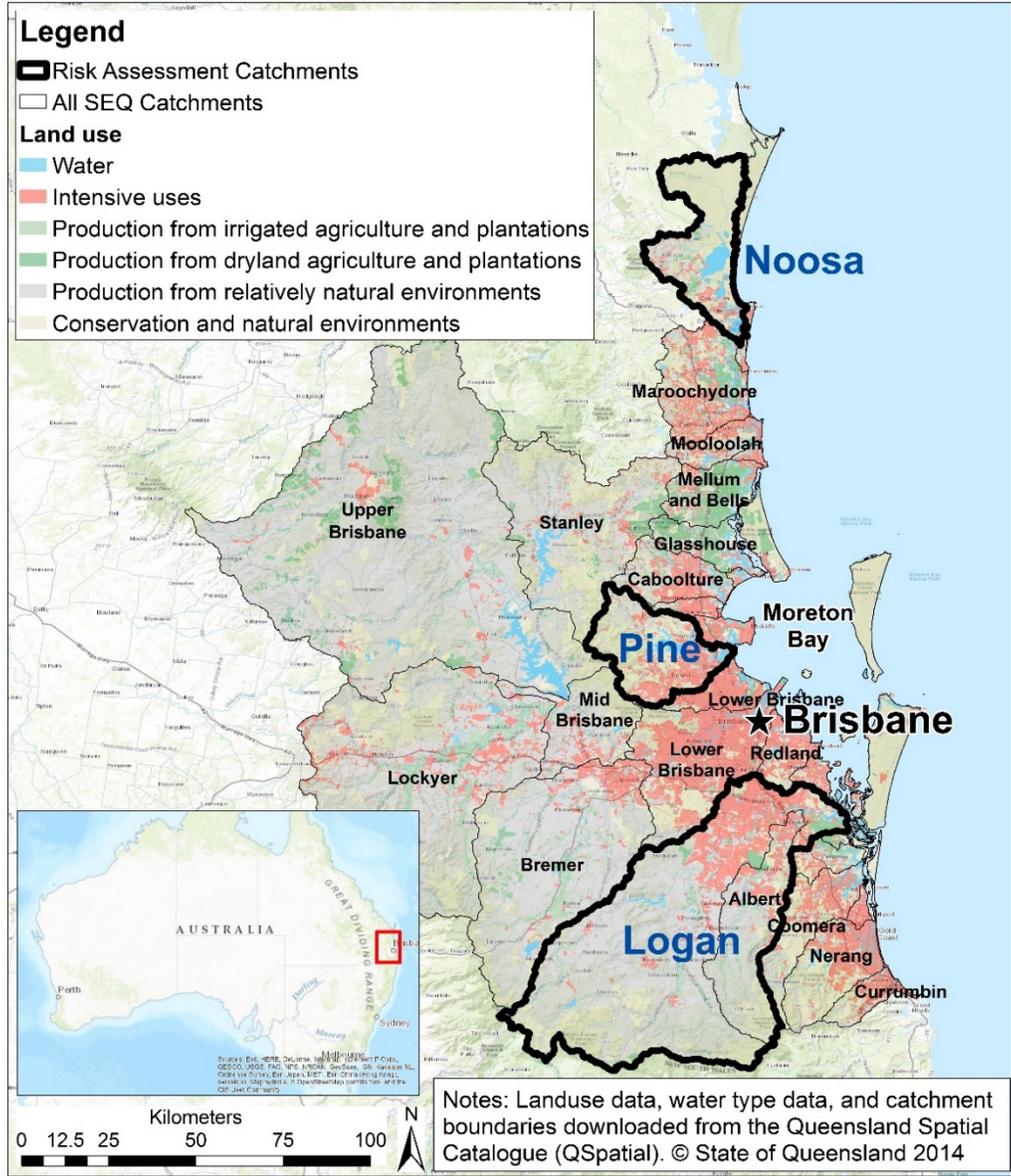
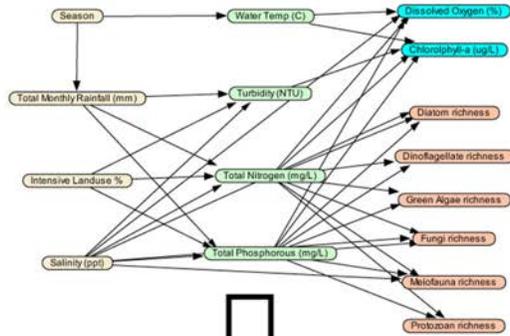


Figure-1. Overview map of the Southeast Queensland region, study sites (Noosa, Pine and Logan catchments), and land use classifications.

Conceptual Model



Logan Estuary Bayesian Network Model

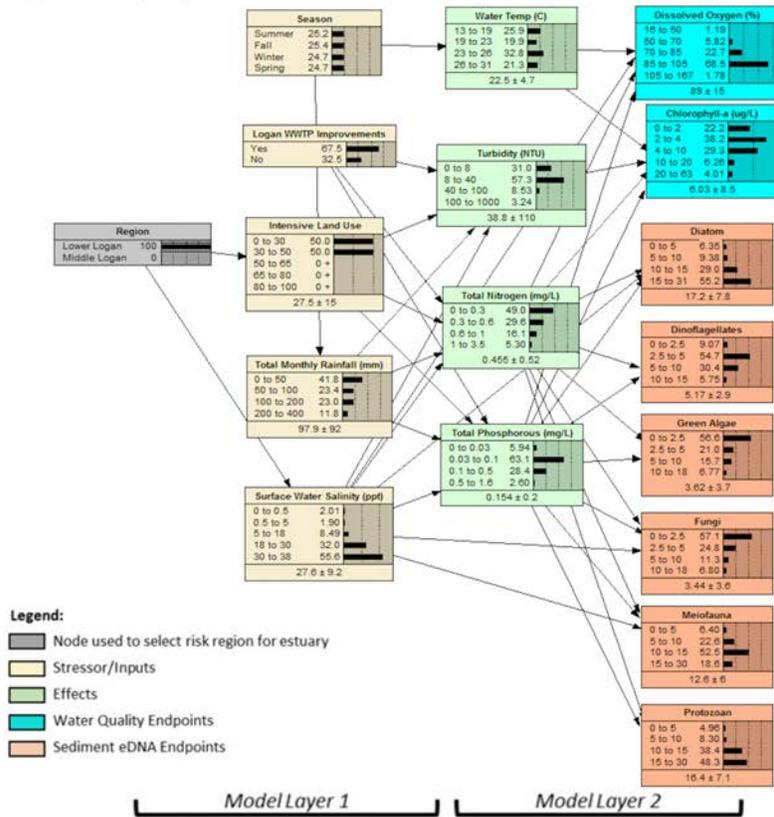
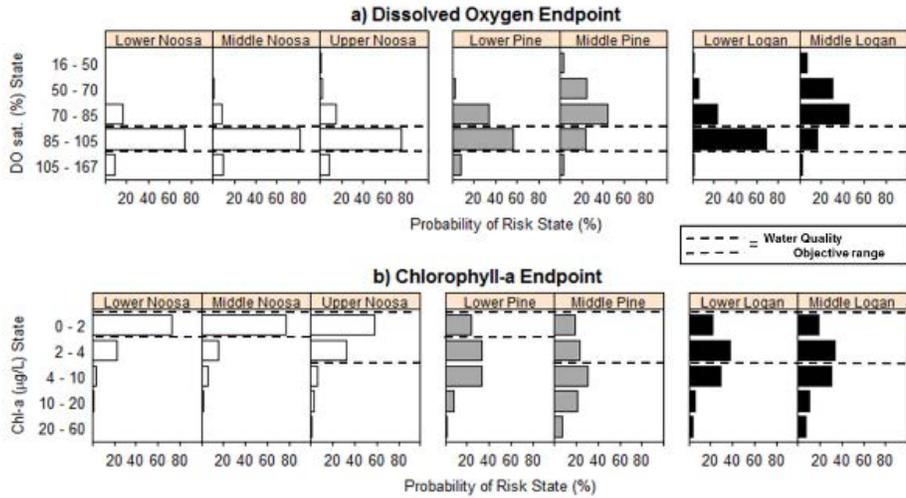


Figure 2. Bayesian network relative risk conceptual model method model and Bayesian network for the Logan estuary with the Middle Logan risk region stressors selected

### Water Quality Endpoints



### Primary Producer Endpoints

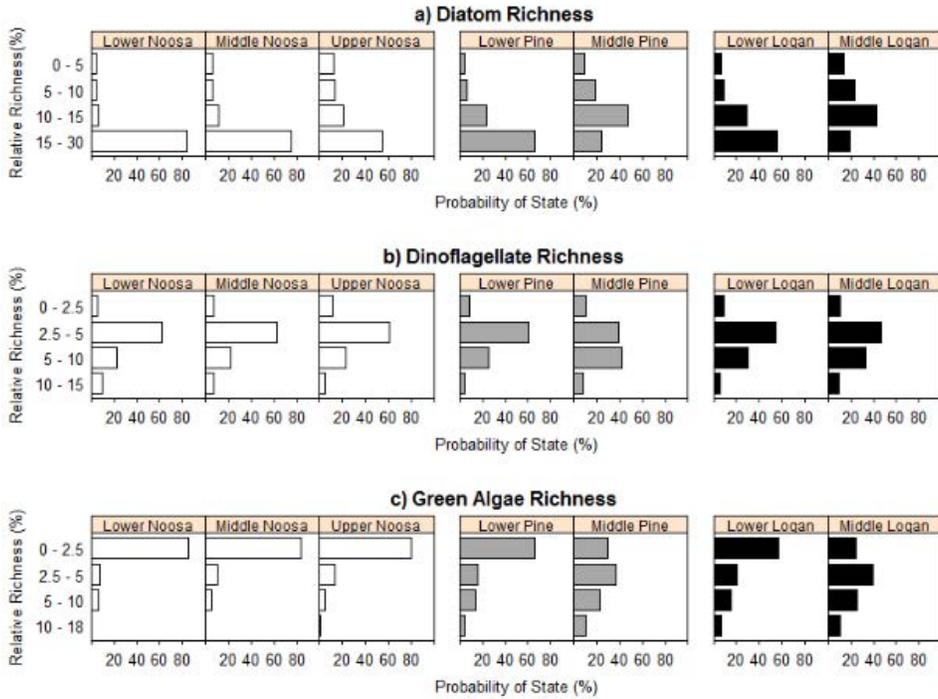
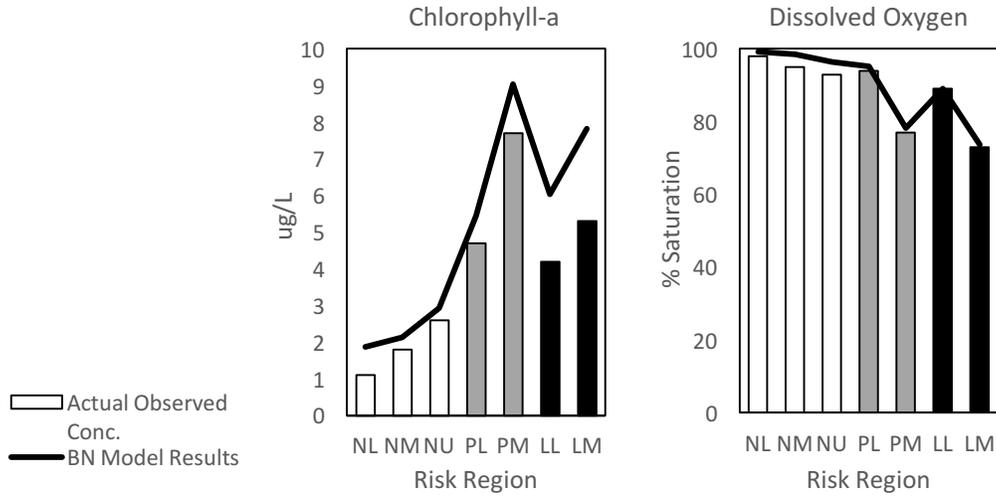


Figure 3- Examples of the output distributions from the BN-RRM calculation: **Water quality and Primary Producers**. The state on the Y-axis labeled “Objective” indicates achieving the regional water quality objective. The objectives and states differ slightly between regions. Also shown are the Primary producer (mostly photosynthetic) posterior probability distributions for each region and estuary. No objectives had been set at the time of this study.



**Figure 4.** Model results for water quality endpoints compared to actual observed average concentrations for each risk region. Abbreviations for the regions are shown in Table 4. The shading of the bars corresponds to the three estuaries: white = Noosa, gray = Pine, and black = Logan.

## Graham et al. Supplemental Information

**Table of Contents**

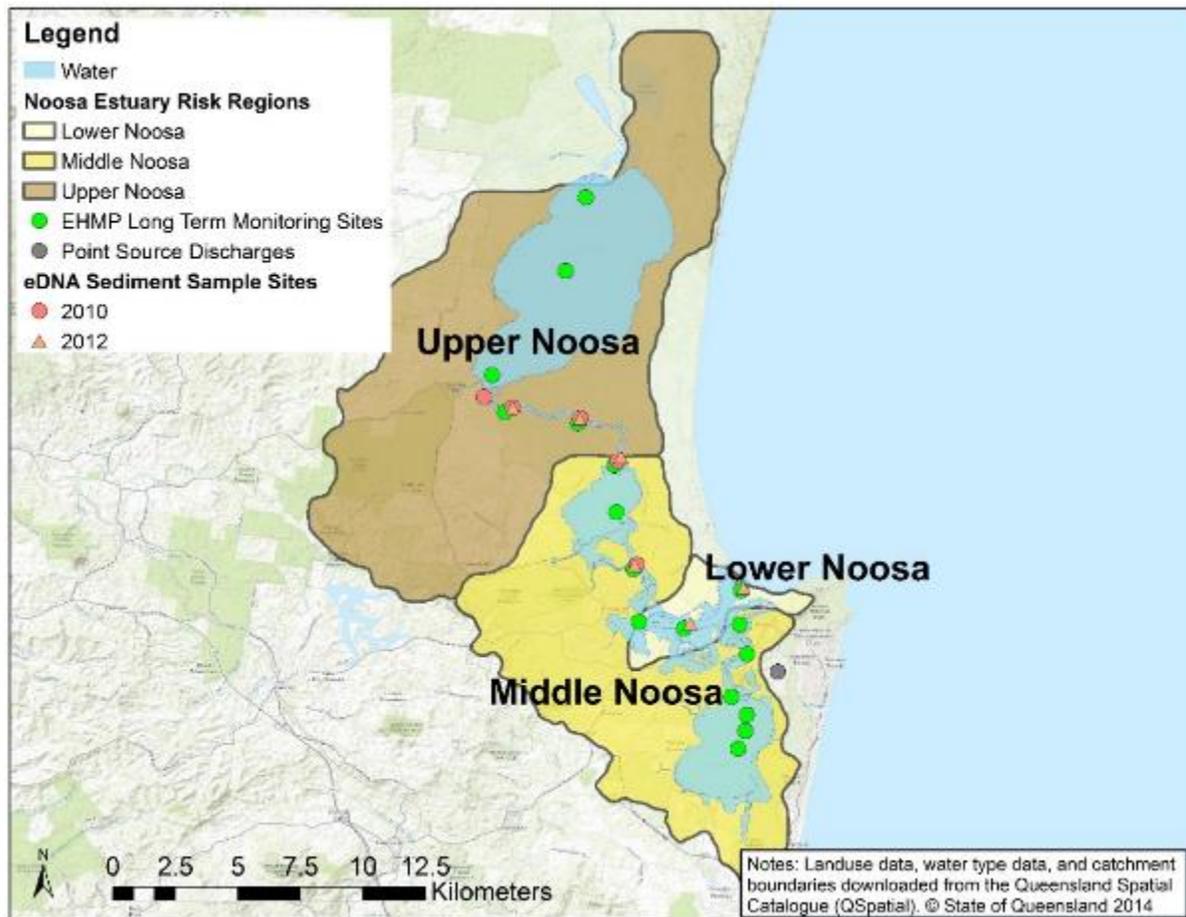
1. Supplemental Tables and Figures.....	2
2. Benthic Environmental DNA sampling and analyses.....	11
3. Sources of information for case learning files.....	23
4. Images of the Bayesian Network models.....	24

## 1. Supplemental Tables and Figures

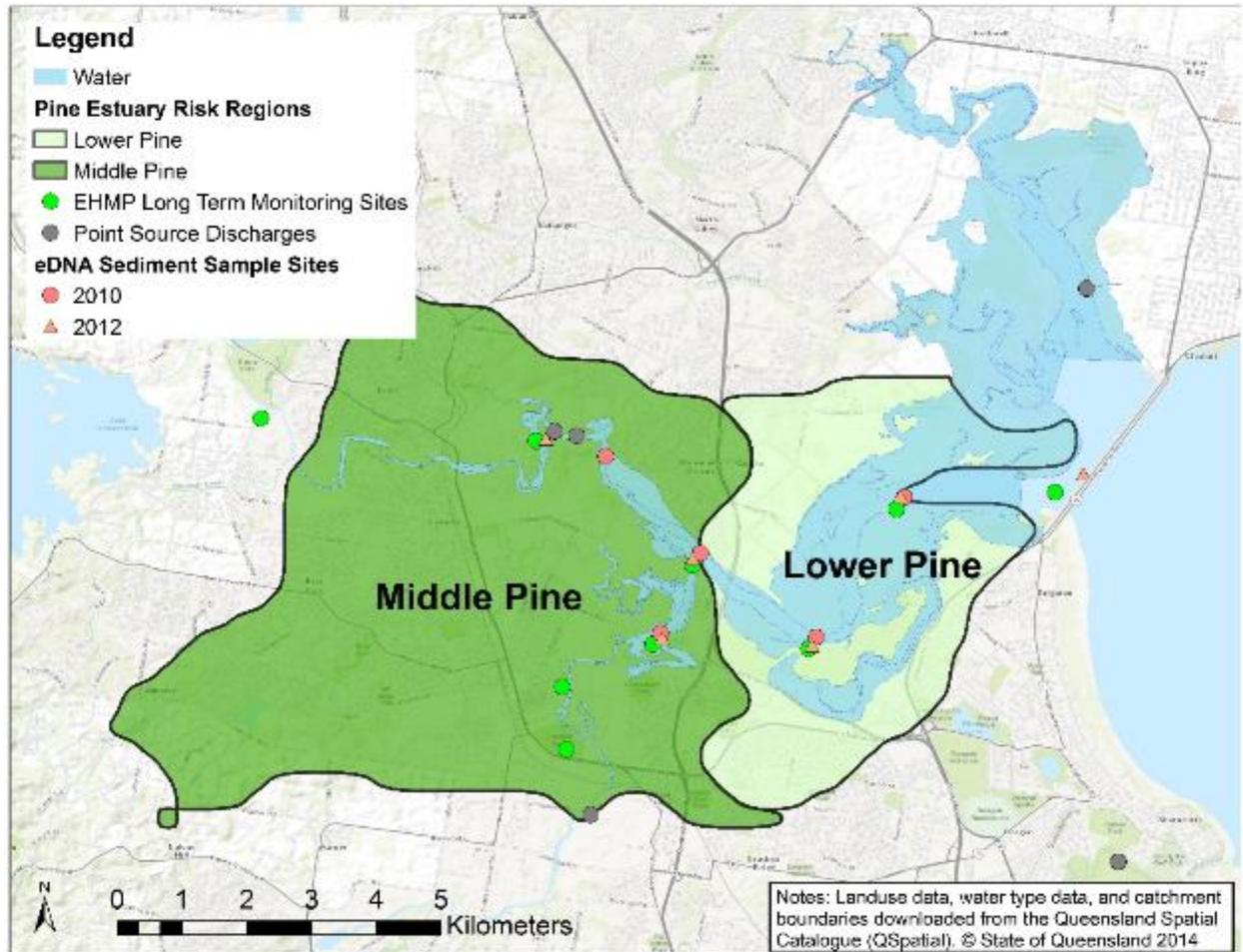
**Table S1.** Validation results for the water quality and benthic taxa endpoints are shown.

Water Quality Endpoints	Cross-validated error rate (%)		
	Noosa	Pine	Logan
Dissolved oxygen	32.47	42.27	31.18
Chlorophyll-a	32.66	56.13	59.13
Mean error rate across all models (standard error)	32.6 (0.1)	49.2 (6.9)	43.2 (14)
Benthic Taxa Endpoints	All estuaries		
Diatom	28.07		
Dinoflagellates	40.35		
Green Algae	40.35		
Meiofauna	35.09		
Fungi	29.82		
Protozoan	33.33		
Mean error rate across all models (standard error)	34.5 (2.1)		

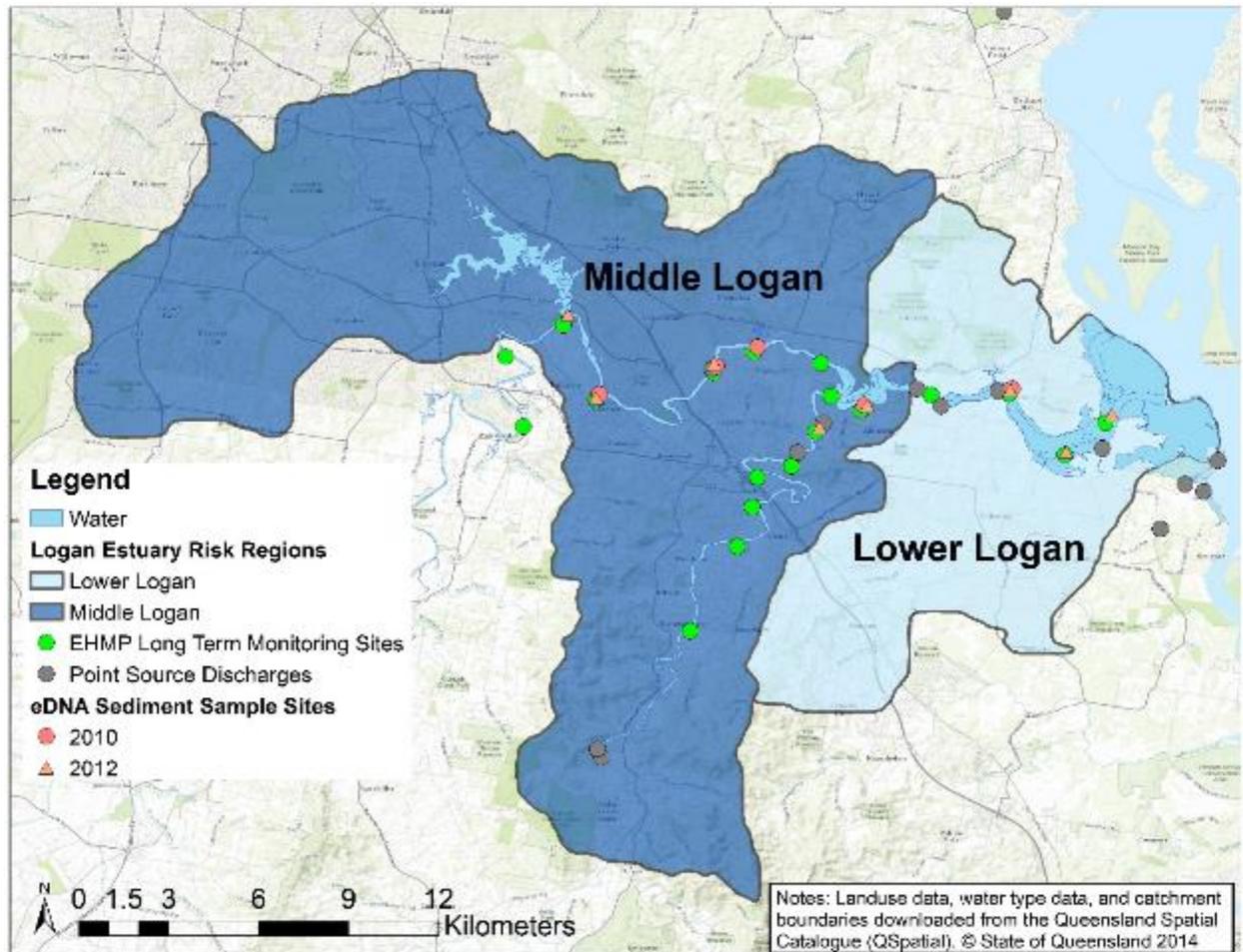
## FIGURES



**Figure S1:** Noosa estuary sub risk regions: Lower Noosa, Middle Noosa and Upper Noosa.

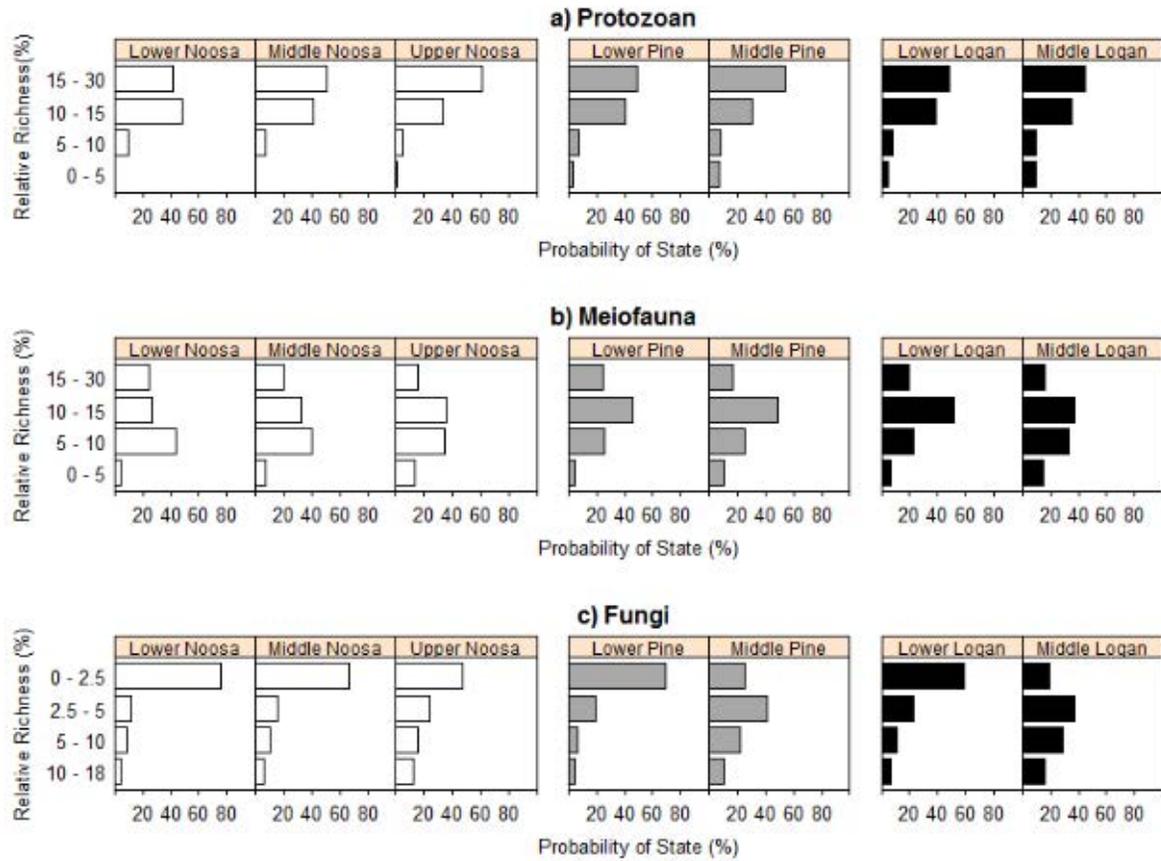


**Figure S2:** Pine estuary sub risk regions: Lower Pine and Middle Pine.

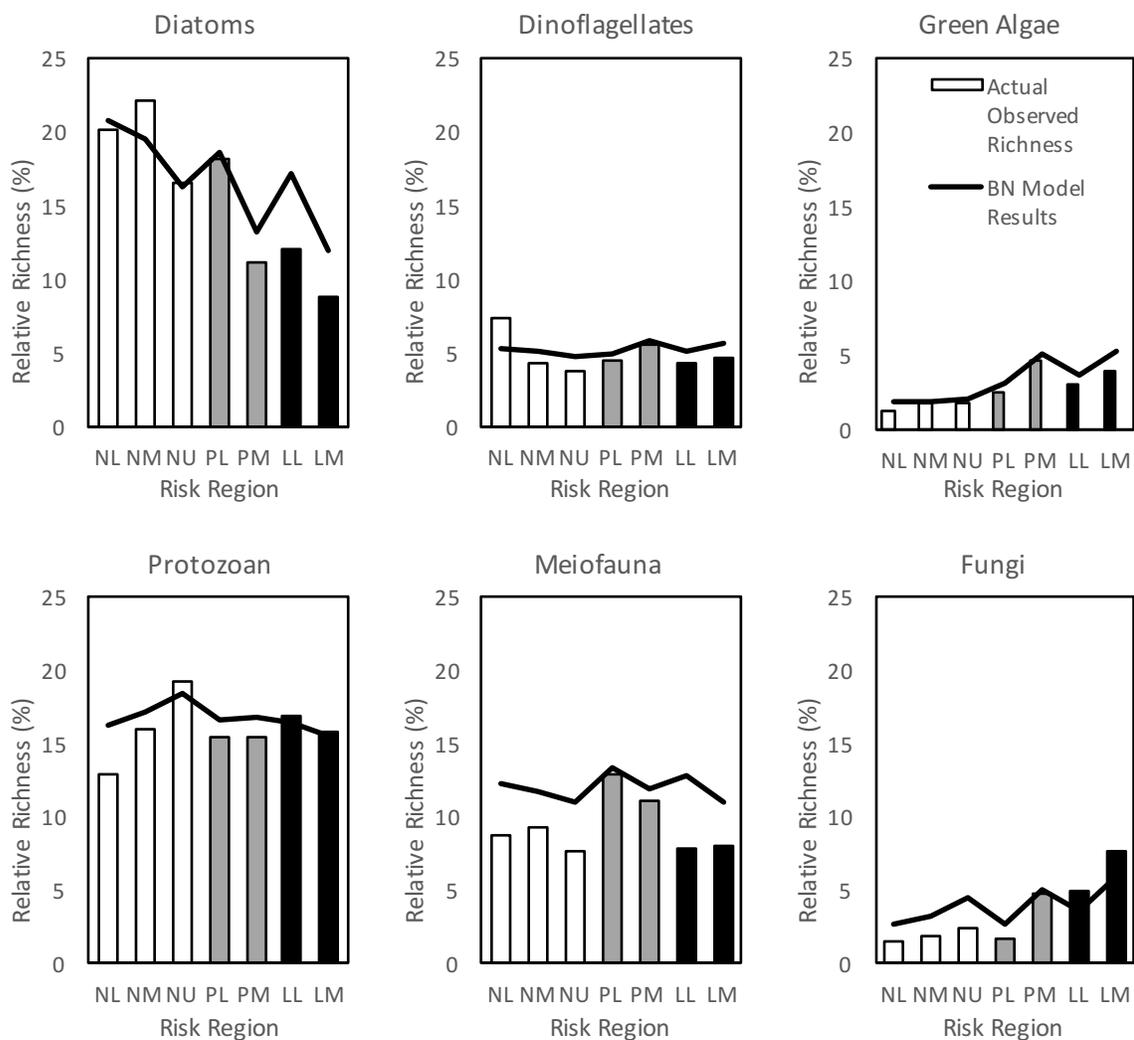


**Figure S3:** Logan estuary sub risk regions: Lower Logan and Middle Logan

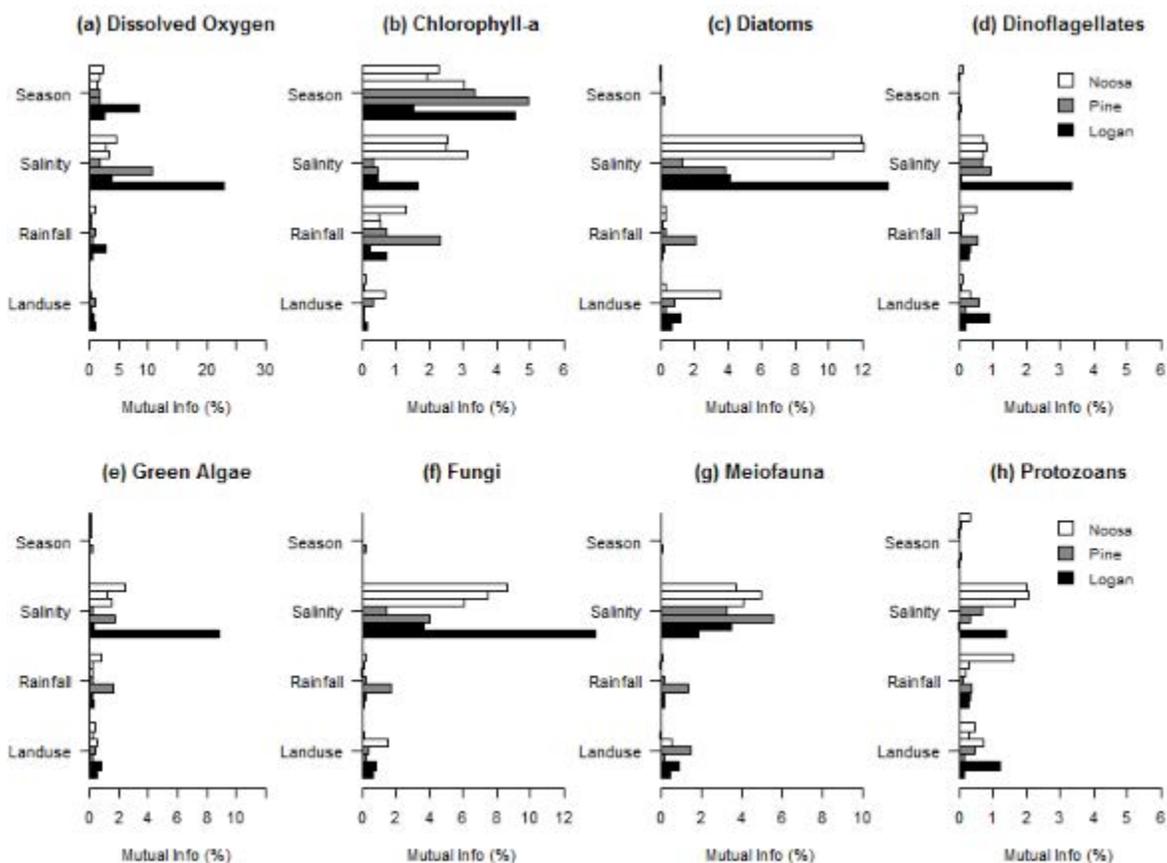
**Primary consumers and decomposer (non-photosynthetic)**



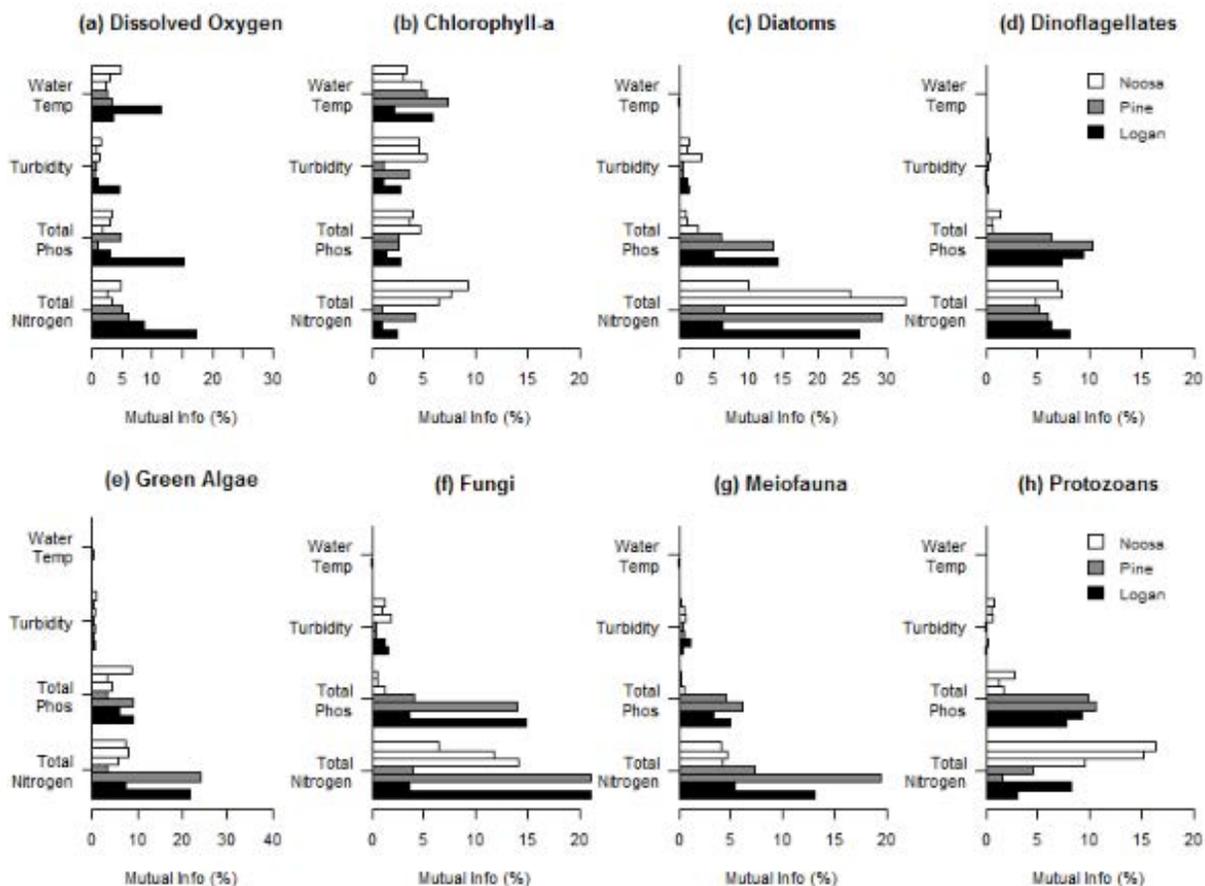
**Figure S4.** Primary consumers and decomposer (non-photosynthetic) posterior probability distributions for each risk region.



**Figure S5:** Model results for relative richness of benthic taxa groups compared to actual observed richness patterns for each risk region. Abbreviations for the region are shown in Table 4. The shading of the bars corresponds to the three estuaries: white = Noosa, gray = Pine, and black = Logan.



**Figure S6.** Mutual info percent from sensitivity analysis for the endpoints to the four input variables (Season, Salinity, Rainfall and Landuse). Results are shown for all regions. Please note different x-axis scale for each endpoint.



**Figure S7.** Mutual info percent from sensitivity analysis for the endpoints to the four intermediate water quality variables (Water Temperature, Turbidity, Total Phosphorus and Total Nitrogen) . Results are shown for all regions. Please note different x-axis scale for each endpoint.

## 2. Benthic environmental DNA sampling and analyses

### *Introduction*

Scientists from CSIRO, Australia's National Science Agency, collected the field samples and sequenced the benthic eDNA data used in this risk assessment. The estuarine surface sediment benthic eDNA samples were collected by CSIRO during two sampling events (Summer 2010 and Fall 2012) from five estuaries in SEQ: the Noosa, Maroochydore, Pine, Logan and Currumbin (Figure S1). While only three estuaries (the Noosa, Pine, and Logan) were evaluated in the risk assessment presented in this thesis, eDNA from all five sampled estuaries were used for determining the water quality predictors in the model and for parameterizing the CPTs.

This section summarizes CSIRO's field collection and laboratory analysis, and my data analysis used to make decisions to incorporate eDNA into the BN-RRM model. The objectives for my analysis of the eDNA data were threefold: 1) to understand patterns in the eDNA benthic communities; 2) to organize the eDNA data into endpoints useful for a risk assessment; and 3) determine linkages in the model between eutrophication water quality variables (TN, TP, Turbidity), Salinity, and benthic community richness.

### *CSIRO field sampling and eDNA analysis*

The methods and results of the 2010 benthic community eDNA investigation have been reported by Chariton et al. (2015), and should be referenced for specific details of the field and analytical methods. This is the first time data from the 2012 sampling event has been presented. The methods of the 2012 sampling event and lab analysis were nearly identical to the 2010 event, as described below.

During field collection, surface sediment grab samples were collected from non-sandy substrates at estuarine sites that were co-located with the EHMP long-term water quality monitoring sites (Table 2.1). Between five to eight sediment grabs were collected from each

estuary per sample event. Sub samples were collected from the grabs for eDNA sequencing, grain size, and total organic carbon analysis. The physio-chemical properties of the water column were measured at each sampling site at a depth of approximately 0.5m above the sediment surface. In addition, surface water grab samples were collected from the same depth as the physio-chemical measurements and analyzed in the laboratory for nutrients (TP, filterable reactive phosphorus, TN, organic nitrogen, inorganic nitrates and ammonia) and Chl-*a* (Chariton et al. 2015). As noted previously, the 2010 and 2012 field and laboratory methods were comparable, however, in 2012, additional sites near the mouth of each estuary were sampled (see **Figures S1-S3**).

**Table 2.1:** Lookup table for the co-located EHMP water quality sites and eDNA sample sites in the Logan, Pine, and Noosa estuary risk regions.

Risk Region	EHMP estuary sites	2010 eDNA sites	2012 eDNA sites
Lower Logan	200		LL1
	201		LL2
	211		
	212	L1	LL3
Middle Logan	202	L2	LL4
	203		
	204	L3	
	205	L4	LL5
	206	L5	LL6
	207		LL7
Lower Pine	800		PP1
	801	P1	PP2
	802	P2	PP3
Middle Pine	803	P4	PP5
	804	P5	PP6
	811	P3	PP4
	812		
	814		
Lower Noosa	1601		NN1
	1603		NN2
Middle Noosa	1604		
	1611		
	1613		
	1614	N2	
	1616		
	1617		
	1624		
	1625		
	1626		
	1636	N1	NN3
Upper Noosa	1608	N3	NN5
	1609	N5	
	1610		
	1615	N4	NN6
	1618		

During laboratory analyses, the eDNA was extracted and purified from five replicates of the sediment samples (Chariton et al. 2015). Three internal reference samples containing clones from multiple eukaryotic taxa groups were also processed. Polymerase chain reaction (PCR) amplification of a 200-500-bp fragment of the 18S rRNA gene was carried out with the universal primers All18SF-TGGTGCATGGCCGTTCTTAGT and All18SR-CATCTAAGGGCATCACAGACC (Chariton et al. 2015). Pyrosequencing was performed by the Australian Genome Research Facility (St Lucia, Queensland) using a single plate of the Roche 454 GFLX Titanium. Cleanup of the sequences including removal of potential PCR artifacts, analogs or multiples of a sequence, errors and chimeras sequences was performed using the Amplicon Pyrosequence Denoising Program (APDP) 20 (Chariton et al. 2015). Taxon identification of the unique sequences, which are referred to as a Molecular Operational Taxonomic Units (MOTU), was inferred using the RDP classifier with the SILVA 18S rRNA database (release 113) ([www.arb-silva.de/](http://www.arb-silva.de/)) (Chariton et al. 2015).

#### *Data analysis: big picture patterns in eDNA -Ordination*

All MOTU data was transformed to presence/absence prior to computation and analysis (Chariton et al. 2014; Chariton et al. 2015), because there is a weak statistical relationship between the number of sequence reads and organism biomass (Egge et al. 2013). After the transformation, we performed multivariate analyses of the 2010 and 2012 MOTUs datasets separately to understand similarities and differences between estuaries during the two sampling events. Analyses included non-metric multidimensional scaling (NDMS) of the MOTUs using the R package *Vegan*.

#### *Incorporating eDNA into the BN-RRM model as endpoints*

To incorporate eDNA information in the BN model, we normalized the 2010 and 2012 samples into a single dataset by classifying the MOTUs into 14 taxa groups based on their

phylum classification (Table 2.2). The taxa group with the highest relative richness for both the 2010 and 2012 samples was the Unclassified organisms, followed by the Protozoans, Diatoms and Meiofauna. For the benthic endpoints, we selected the six taxa groups with the highest richness in both 2010 and 2012, excluding the Unclassified group. Definitions of how these six groups were sorted based on taxonomic information from the SILVA database are provided in Table 2.3. Correlations were performed between relative richness of these groups and water quality variables to further determine predictors in the model.

**Table 2.2:** Summary of average relative richness for taxa groups. Groups shown in bold and with a \* were used in the risk assessment because they had the highest richness across both sampling events.**Error! Not a valid link.**

**Table 2.3:** Taxonomic definitions of the six most taxa groups used in the risk assessment.**Error! Not a valid link.**

*Determining water quality predictors for the eDNA endpoints*

A combination of methods and information was used to determine the predictors and linkages in the BN-RRM model including vector fitting, the BEST test, simple correlations, and via trial and error of comparing predicted richness to observed richness in all risk regions.

CSIRO water quality measurements of TN, TP, Turbidity and pH were compared to EHMP measurements matched based on the closest sampling dates. The EHMP data was comparable to the CSIRO data, and was ultimately used in determining predictor variables for the BN model, and also for CPT parameterization. While the CSIRO nutrient data is more representative of the concentrations experienced by the benthic organisms, the EHMP data is preferred for the model because it was collected in the same manner as the rest of data used in the model. All of the nutrient samples from the EHMP data were collected and analyzed via the same method and same depth (0.2m below the water surface). The CSIRO samples on the other hand were collected at variable depths (approximately 0.5m above the sediment surface). In addition, there were some missing samples in the CSIRO data, so using the more complete EHMP dataset made for more complete case-files for the CPT parameterization. One final reason for use of the EHMP water quality data was that the CSIRO samples did not include salinity measurements which were needed for the model.

For vector fitting, the centroids of the five benthic sample replicates per site were used to fit the water quality variables. Chariton et al. (2015) also calculated centroids to handle the replicates when fitting to environmental variables. Vector fitting was performed with the `envfit` function also in the R package, `Vegan`, and related NMDS patterns based on the centroids to the water quality variables. We only looked at the water quality variables that

were included in the BN-RRM model (TN, TP, Turbidity, Salinity, DO, Chl-a, and Temperature) because these were previously determined to be the most important direct and indirect effects of eutrophication in SEQ.

The BEST test was performed using the bioenv function from the Vegan package. The BEST test compares the biotic richness of one of the taxa groups with an array of abiotic matrices formed by the water quality predictor variables (Clarke et al. 2008; Lucena-Moya et al. 2015). From this, the variables that best explained the taxa group richness were identified and evaluated for inclusion as predictors in the final BN-RRM. The variables TN, TP, Turbidity, and Salinity were evaluated using the BEST test. Temperature was initially included and results indicated that it was a good predictor of some groups. However, when we included it as a predictor in the BN-RRM model, it did not accurately predict the benthic endpoint richness. So in this case, we re-ran the BEST test looking at only the four variables mentioned above. Furthermore, temperature was not identified by the vector fitting as a good predictor of the MOTUs (results discussed in more detail below).

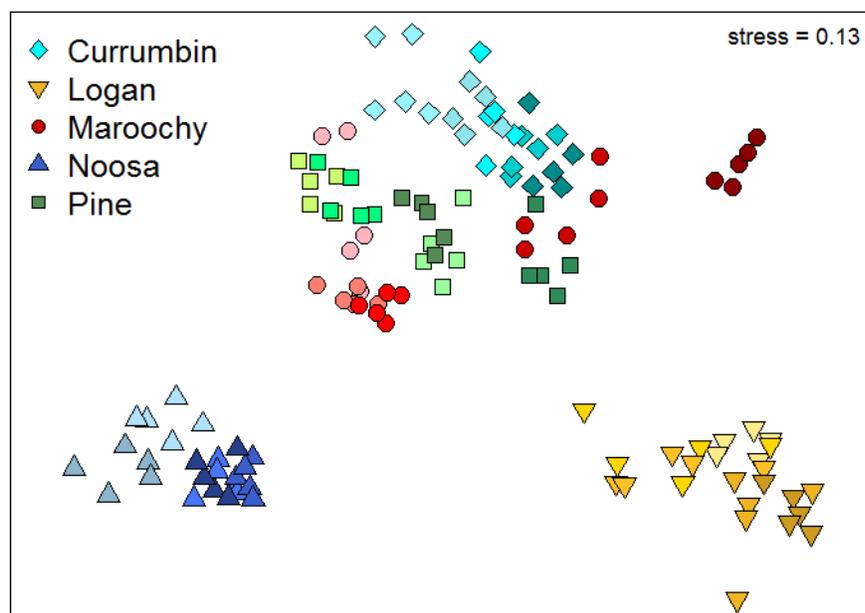
Simple correlations were also performed between EHMP matched monitoring data and taxa group richness to help determine which predictors would be best for the BN-RRM model. The  $R^2$  of the correlations were summarized and used to help determine the predictors in the BN-RRM model.

#### *Big Picture eDNA analysis – NMDS Ordination*

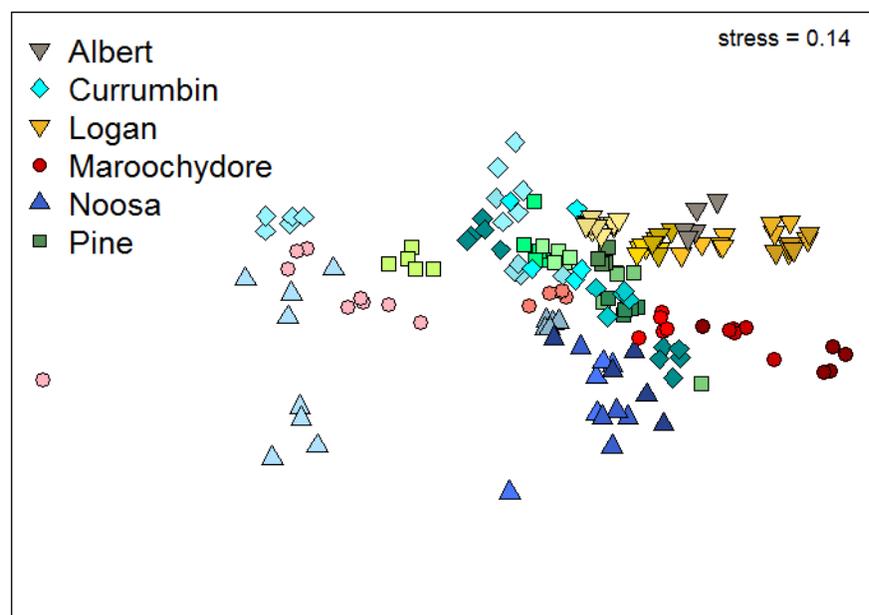
NMDS ordination was conducted using the Jaccard similarity metric for the Summer 2010 and Fall 2012 MOTU data (Figure 2.1). For the 2010 samples, the Noosa and Logan samples separate from the other estuaries in the NMDS plots, indicating that they have different benthic community composition than the other estuaries (Figure 2.1a). The remaining three estuaries (Pine, Currumbin, and Maroochydore) form a third separate group of points, indicating that they are not different from each other. For the 2012 samples, the

samples from the different estuaries do clearly separate via ordination (Figure 2.1b). However, sample points reflecting the same estuary do plot next to each other on the NMDS figures, with the Noosa and Logan again forming the clearest groups. In 2012, the samples from the marine portion of the estuaries with higher salinities generally clustered together on the left side of the plot irrespective of estuary, indicating these samples are similar to each other. The 2010 and 2012 datasets together represent snapshots of the condition of the benthic community composition and reflect the inherent variability of benthic taxa.

## a) 2010 MOTUs



## b) 2012 MOTUs



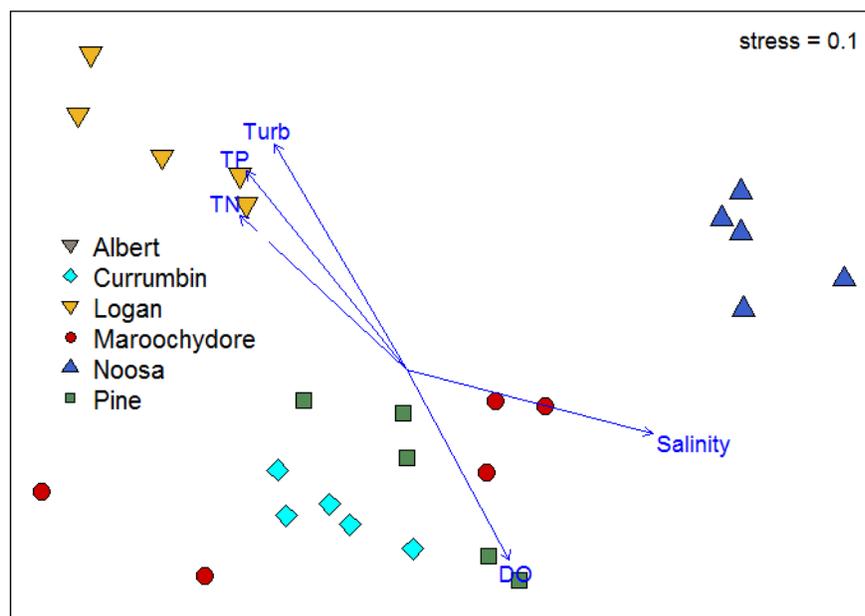
**Figure 2.1.** NMDS plots illustrating the similarities and differences in the benthic eukaryote communities from the five SEQ estuaries for the a) 2010 and b) 2012 sampling events. The shading of the site markers indicates their position from marine (light) to freshwater (dark). Generally, an NMDS plot with stress < 0.2 is considered a good representation of the data.

### *Vector fitting to NMDS ordination*

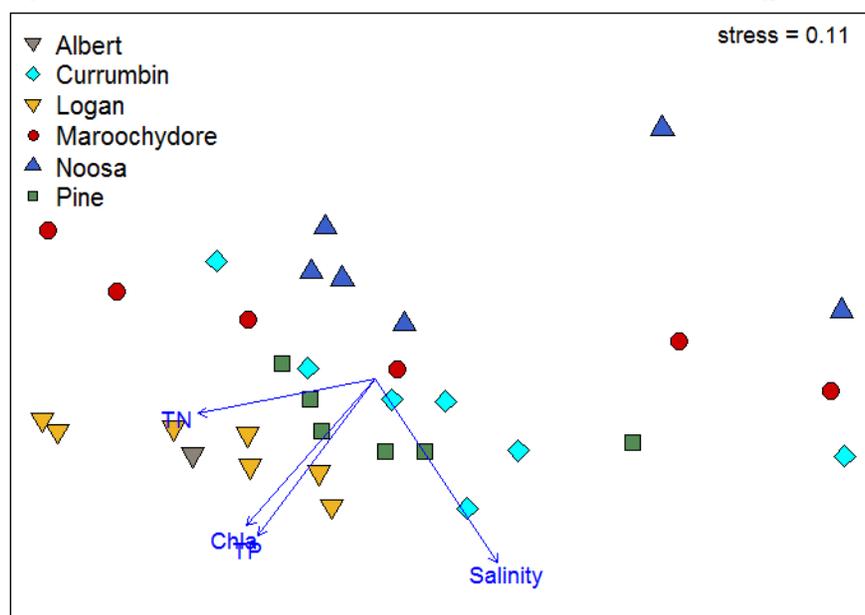
Vector fitting to the NMDS plots was used to visually explain the benthic ordination with the water quality variables, and to begin to understand which variables are consistently most predictive of the benthic taxa (Figure 2.2). In 2010, the water quality variables TN, TP, Turbidity, Salinity, and DO were the best predictors ( $p < 0.001$ ) (Figure 2.2a). The labeled arrows in the vector fitting plots point to the direction of most rapid change in water quality variable, or the direction of the gradient. The length of the arrow is proportional to the correlation between ordination and variable, otherwise known as the strength of the gradient. According to the vector fitting, the TN, TP, and Turbidity gradient is strongest in the Logan, which explains why that estuary clearly separates from the others. The Salinity and DO arrows do not clearly point in the direction of a single estuary, but they do point away from the Logan, indicating the Salinity and DO gradient are most important in the other estuaries compared to the Logan (Figure 2.2a).

In 2012, the water quality variables TN, TP, Chl<sub>a</sub>, and Salinity were the best predictors ( $p < 0.05$ ) (Figure 2.2b). Again, the TN and TP gradient was strongest in the direction of the Logan. The Chl<sub>a</sub> endpoint also appears to be predictive of the Logan. The Salinity gradient, again, does not point in the direction of a specific estuary (Figure 2.2b).

a) 2010 MOTU centroids fit to environmental data ( $p < 0.001$ )



b) 2012 MOTU centroids fit to environmental data ( $p < 0.05$ )



**Figure 2.2.** NMDS plots illustrating the similarities and differences in the benthic eukaryote communities from the five SEQ estuaries for the a) 2010 and b) 2012 sampling events. Water quality variables that are used as predictors in the BN model have been fit to the plot. Generally, an NMDS plot with stress  $< 0.2$  is considered a good representation of the data.

*BEST test and correlations results*

The results of the correlation coefficients and  $R^2$  from the BEST test and correlations, respectively, are provided in Table S4. Based on these results, different predictors were tried in the BN-RRM model. The water quality variables ultimately selected (Table S4, column on the far right) were identified in multiple tests and were confirmed in the BN-RRM model by comparing observed richness to predicted richness. If no variables or only one variable was identified by the different analysis/testing methods, then TN and TP were used as predictors. In the end, TN and TP were used as predictors for all benthic groups. In some cases, prior knowledge from the literature was also used to determine the predictors, and Salinity was selected as a predictor for Diatoms, Meiofauna, and Fungi. Turbidity was not selected as a predictor for any groups.

**Table 2.4:** Summary of analysis used to determine predictors for the BN-RRM model and the final predictors selected.

<b>Taxa Group</b>	<b>Vector Fitting</b> (for MOTUs)	<b>Best Test Results</b> (for taxa group richness)	<b>Correlation Results</b> (for taxa group richness)	<b>Priors from Literature</b> (for taxa group richness)	<b>Predictors used in the BN-RRM model</b>
<b>Diatom</b>	<u>2010</u> MOTUs: <b>TN,</b> <b>TP,</b> Turbidity, <b>Salinity</b>	<b>Salinity, Turbidity</b> (cor = 0.35)	<b>TN (<math>R^2 = 0.25</math>)*</b> <b>TP (<math>R^2 = 0.27</math>)*</b> Turbidity ( $R^2 = 0.16$ ) <b>Salinity (<math>R^2 = 0.35</math>)*</b>	<b>Salinity</b>	<b>TN, TP, and Salinity</b>
<b>Dinoflagellate</b>		Salinity (cor = 0.07)	<b>TN (<math>R^2 = 0.006</math>)</b> <b>TP (<math>R^2 = 0.0004</math>)</b> Turbidity ( $R^2 = 0.001$ ) Salinity ( $R^2 = 0.002$ )		
<b>Green Algae</b>	<u>2012</u> MOTUs: <b>TN,</b> <b>TP,</b> <b>Salinity</b>	<b>TN, Salinity (cor = 0.17)</b>	<b>TN (<math>R^2 = 0.19</math>)*</b> <b>TP (<math>R^2 = 0.032</math>)</b> Turbidity ( $R^2 = 0.022$ ) Salinity ( $R^2 = 0.1$ )		<b>TN, TP</b>
<b>Fungi</b>		<b>Salinity (cor = 0.3)</b>	<b>TN (<math>R^2 = 0.15</math>)</b> <b>TP (<math>R^2 = 0.23</math>)*</b> Turbidity ( $R^2 = 0.14$ ) <b>Salinity (<math>R^2 = 0.29</math>)*</b>		<b>TN, TP, and Salinity</b>
<b>Meiofauna</b>		<b>TN (cor = 0.11)</b>	<b>TN (<math>R^2 = 0.04</math>)</b> <b>TP (<math>R^2 = 0.07</math>)</b> Turbidity ( $R^2 = 0.05$ ) <b>Salinity (<math>R^2 = 0.06</math>)</b>	<b>Salinity</b>	<b>TN, TP, and Salinity</b>
<b>Protozoan</b>		Salinity (cor = 0.08)	<b>TN (<math>R^2 = 0</math>)</b> <b>TP (<math>R^2 = 0.0008</math>)</b> Turbidity ( $R^2 = 0.01$ ) Salinity ( $R^2 = 0.002$ )		<b>TN, TP</b>

### **3. Sources of information for case learning files**

Land Use Data: The land use data for South East Queensland was downloaded free of charge through the Queensland Government GIS dataset webpage.

<https://data.qld.gov.au/dataset/land-use-mapping-series>

Precipitation Data: Precipitation data was downloaded from Queensland BOM weather stations free of charge.

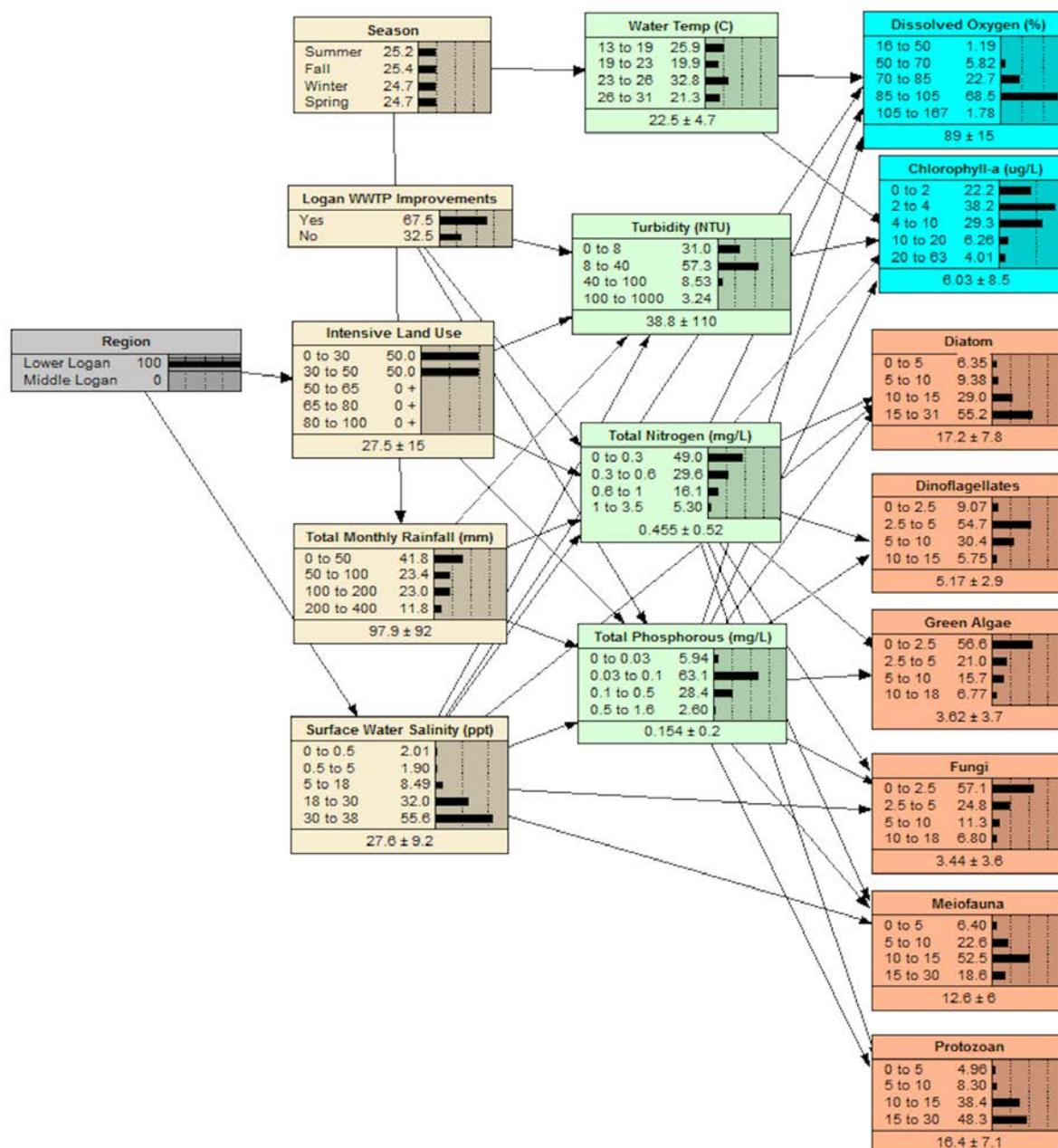
Water Quality Monitoring Data: The water quality data was provided free of charge through data requests to the South East Queensland Health-e-waterways program. <http://hlw.org.au/report-card/monitoring-program>.

eDNA Data: The benthic eDNA dataset was provided by CSIRO as described in Section 2 of this Supplement Section. ([www.arb-silva.de/](http://www.arb-silva.de/))

## 4. Images of the Netica-BN-RRM networks

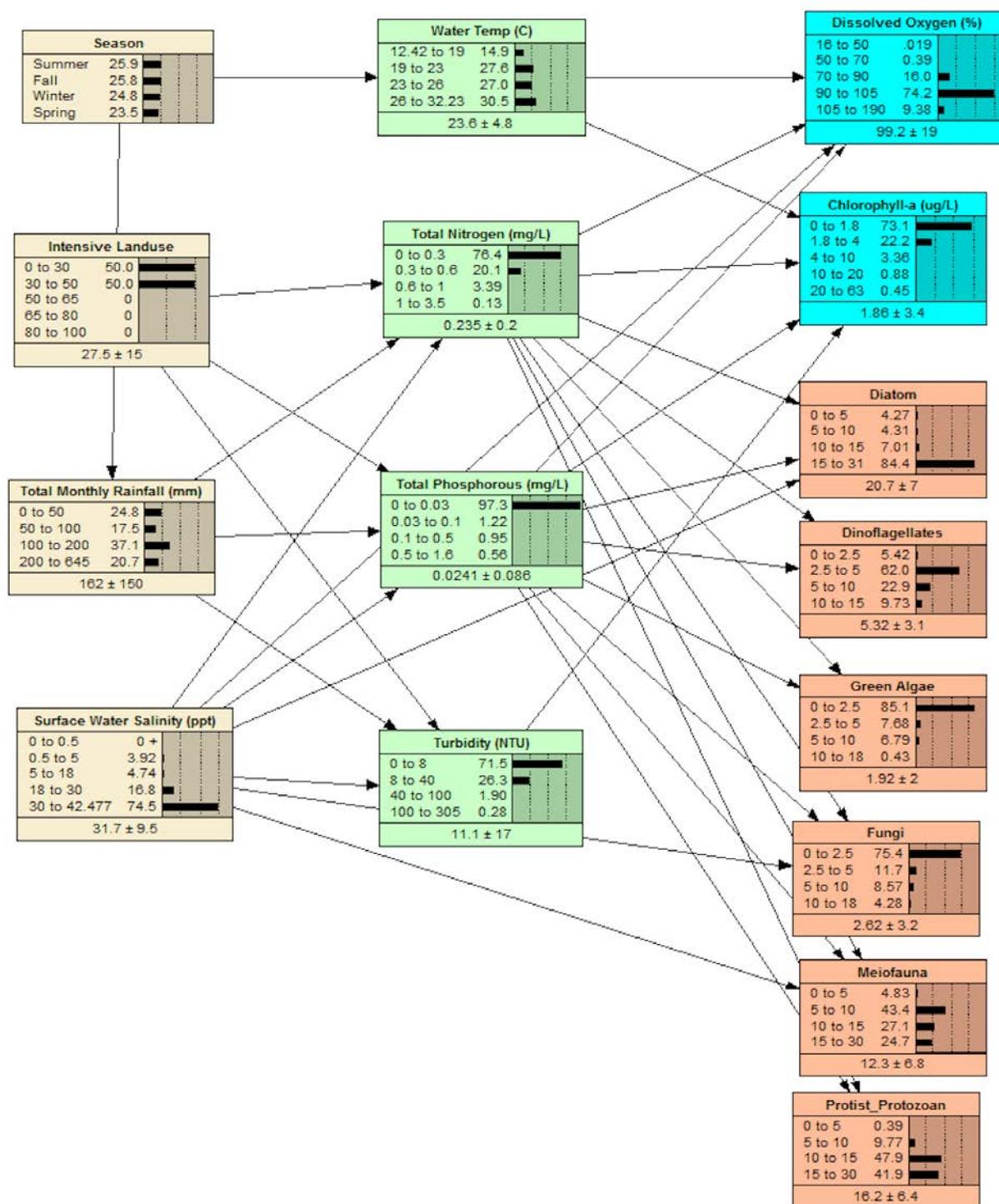
The names of the models correspond to the Netica file names also included as supplemental information.

### Lower Logan



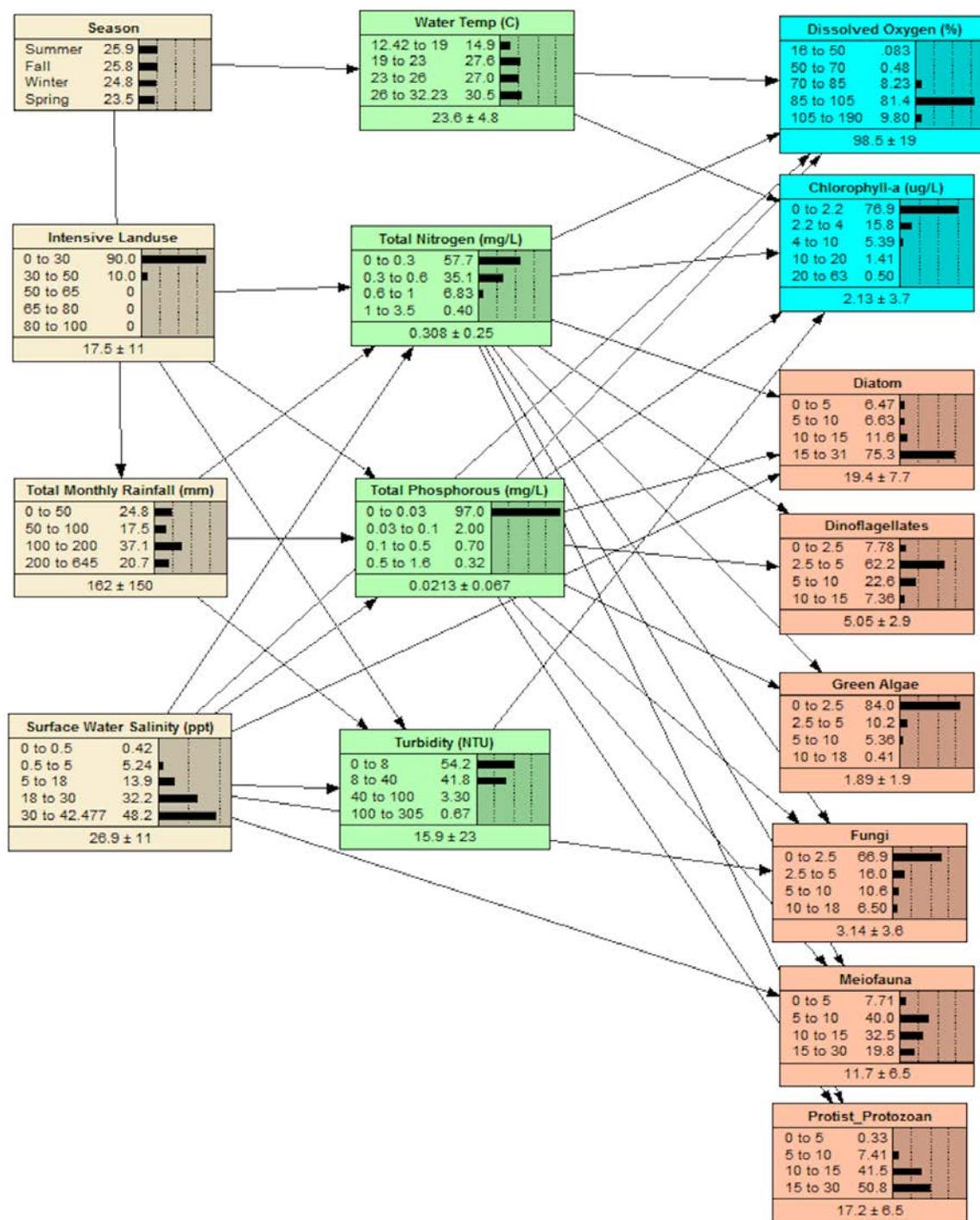
**Figure 4.1:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan estuary with the **Lower Logan** risk region stressors selected. See the LOGAN.neta for the Netica model file.

## Lower Noosa



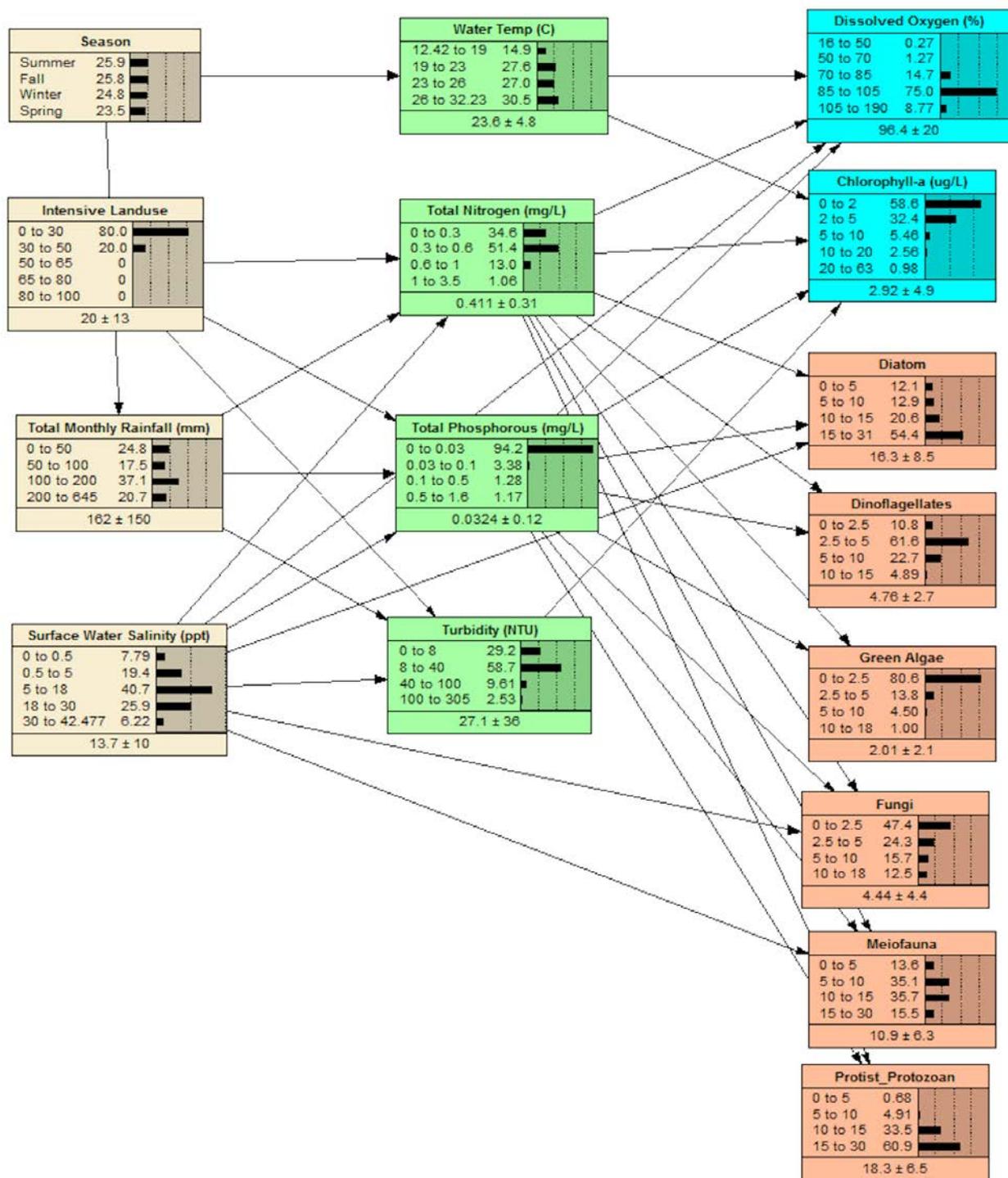
**Figure 4.2:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan estuary with the **Lower Noosa** risk region stressors selected. See the NOOSA\_LOWER.neta for the Netica model file.

## Middle Noosa



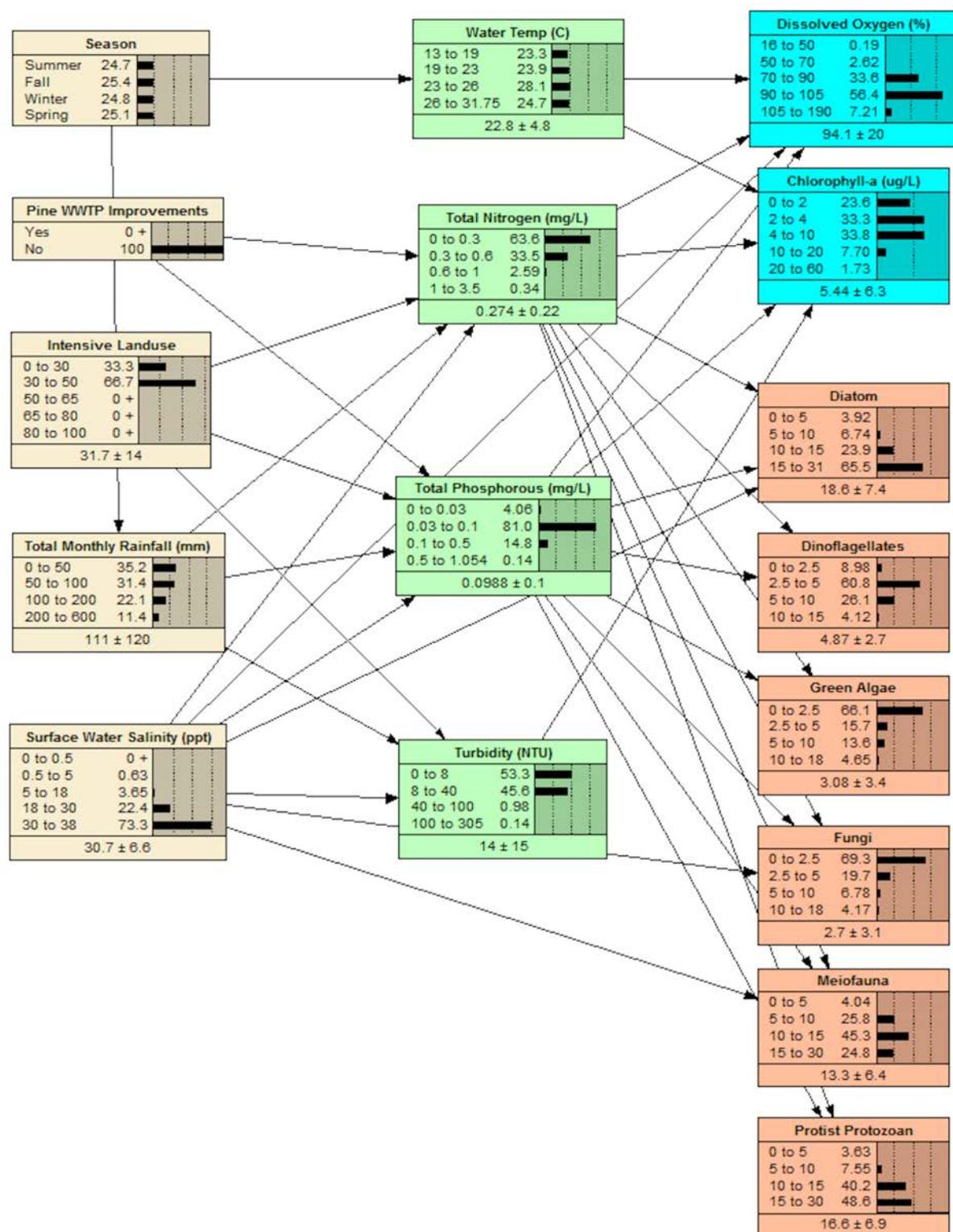
**Figure 4.3:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan estuary with the **Middle Noosa** risk region stressors selected. See NOOSA\_MIDDLE.neta for the Netica model file.

## Upper Noosa



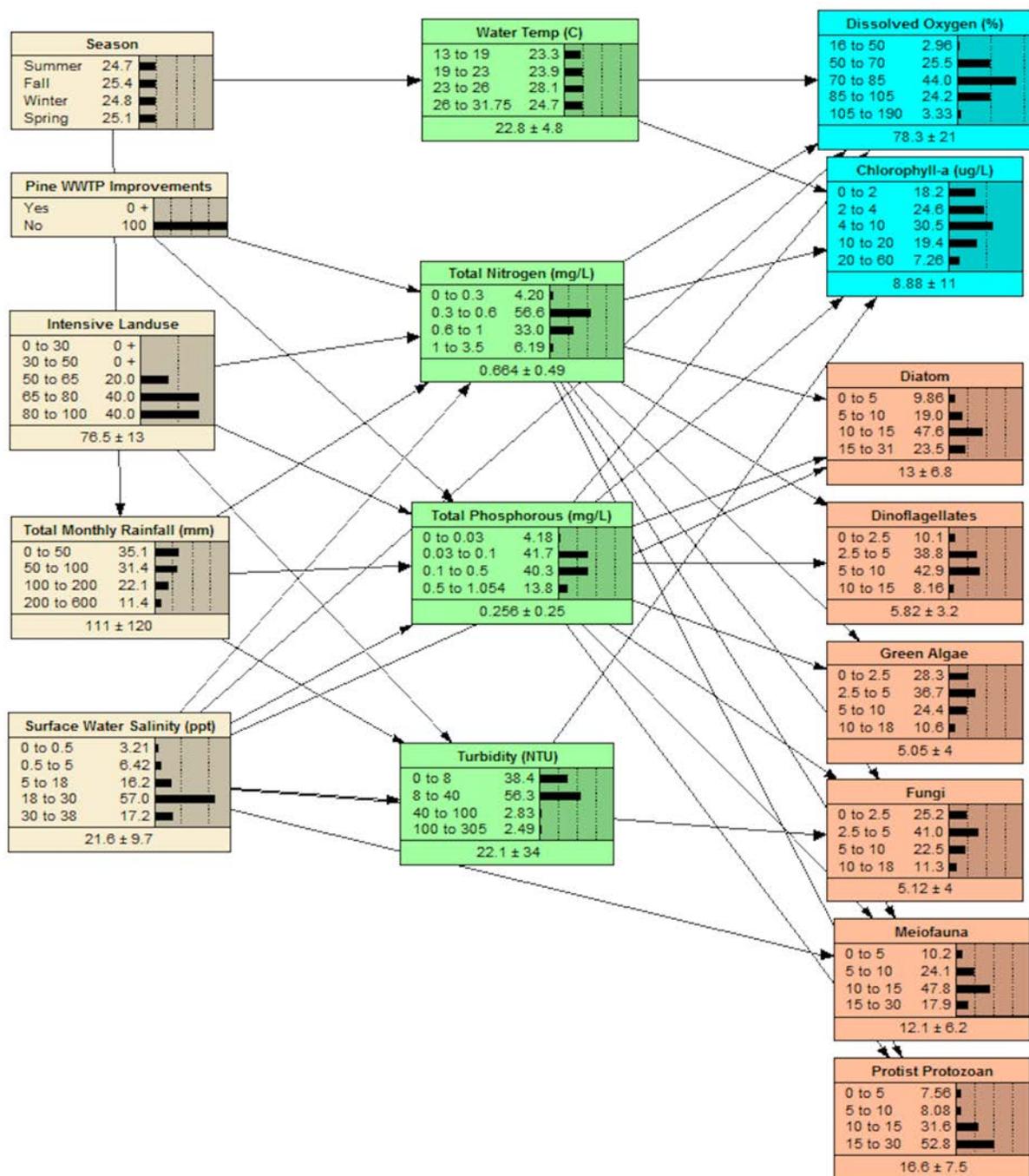
**Figure 4.4:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan estuary with the **Upper Noosa** risk region stressors selected. See the NOOSA\_UPPER.neta for the Netica model file.

## Lower Pine



**Figure 4.5:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan estuary with the **Lower Pine** risk region stressors selected. See PINE\_LOWER.neta for the Netica model file.

## Middle Pine



**Figure 4.6:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan estuary with the **Middle Pine** risk region stressors selected. See PINE\_MIDDLE.neta for the Netica model file.