Spring 2016

Predicting Risk to Estuary Water Quality and Patterns of Benthic Environmental DNA in Queensland, Australia using Bayesian Networks

Scarlett E. Graham
Western Washington University, segraham85@gmail.com

Follow this and additional works at: https://cedar.wwu.edu/wwuet

Part of the Environmental Sciences Commons

Recommended Citation
Graham, Scarlett E., "Predicting Risk to Estuary Water Quality and Patterns of Benthic Environmental DNA in Queensland, Australia using Bayesian Networks" (2016). WWU Graduate School Collection. 511.
https://cedar.wwu.edu/wwuet/511

This Masters Thesis is brought to you for free and open access by the WWU Graduate and Undergraduate Scholarship at Western CEDAR. It has been accepted for inclusion in WWU Graduate School Collection by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.
Predicting Risk to Estuary Water Quality and Patterns of
Benthic Environmental DNA in Queensland, Australia using Bayesian Networks

By
Scarlett E. Graham

Accepted in Partial Completion
Of the Requirements for the Degree
Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Wayne G. Landis

Dr. Anthony A. Chariton

Dr. Robin Matthews

Dr. David H. Shull
MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Scarlett E. Graham

20 May 2016
Predicting Risk to Estuary Water Quality and Patterns of Benthic Environmental DNA in Queensland, Australia using Bayesian Networks

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By

Scarlett E. Graham

20 May 2016
ABSTRACT

Predictive modeling can inform natural resource management by demonstrating stressor-response pathways and quantifying the effects on selected endpoints. This study develops a risk assessment model using the Bayesian network-relative risk model (BN-RRM) approach, and, for the first time, incorporates eukaryote environmental DNA data as a measure of benthic community structure into an ecological risk assessment context. Environmental DNA sampling is a relatively new technique for biodiversity measurements that involves extracting DNA from environmental samples, sequencing a region of the 18s rDNA gene, and matching the sequences to organisms. Using a network of probability distributions, the BN-RRM model predicts risk to water quality objectives and also the richness of benthic taxa in the Noosa, Pine, and Logan Estuaries in South East Queensland (SEQ), Australia. The model is more accurate at predicting Dissolved Oxygen than it is the Chlorophyll-a water quality endpoint, and it predicts photosynthesizing benthos more accurately than heterotrophs. Results of BN-RRM modeling 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 high probability (73 - 92% probability) of achieving water quality objectives, which indicates low relative risk. On the other hand, the Middle Logan, Middle Pine, and Lower Pine regions are much less likely to meet objectives (15 – 55% probability), indicating a relatively high risk to water quality in those regions. The benthic community richness patterns associated with low relative risk in the Noosa are high Diatom relative richness and low Green Algae richness. The only benthic pattern consistently associated with high relative risk to water quality is the high Fungi richness state. The BN-RRM predicts current conditions in SEQ based on available monitoring data, and provides a basis for future predictions and adaptive management at the direction of resource managers. As new data are made available or more questions are asked, this BN-RRM model can be updated and improved.
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 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. The Winter 2016 ESCI 597B Scientific Writing class held at Shannon Point Marine Center with Sylvia Yang was also a godsend during the writing process, providing much guidance and inspiration.

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.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................ iv
Acknowledgements ...................................................................................................................... v
List of Figures ........................................................................................................................... ix
List of Tables .............................................................................................................................. ix
List of Supplemental Information ............................................................................................ x
Acronyms and Abbreviations .................................................................................................... xi

1. INTRODUCTION .................................................................................................................. 1
   1.1 South East Queensland ......................................................................................................... 1
   1.2 Ecological Risk Assessment and the Relative Risk Model (RRM) ........................................ 2
   1.3 Bayesian Network (BN) Modeling ..................................................................................... 3
   1.4 Environmental DNA (eDNA) ............................................................................................ 4
   1.5 Study Objectives .............................................................................................................. 5
   1.6 Summary of Findings ........................................................................................................ 5

2. SOUTH EAST QUEENSLAND (SEQ) STUDY REGION ....................................................... 7
   2.1 Climate ............................................................................................................................. 8
   2.2 Land Use ........................................................................................................................ 10
   2.3 Management of SEQ Waterways ...................................................................................... 11
   2.4 Noosa ............................................................................................................................. 12
   2.5 Pine ................................................................................................................................ 14
   2.6 Logan .............................................................................................................................. 16

3. METHODS ........................................................................................................................... 18
   3.1 Relative Risk Model (RRM) Overview ............................................................................. 18
      3.1.1 Selection of Endpoints ............................................................................................. 19
      3.1.2 Identification of Stressors and Sources of Stressors ............................................... 22
      3.1.3 Risk Region Delineation ......................................................................................... 23
      3.1.4 Model Structure ...................................................................................................... 23
      3.1.5 Model Assumptions ............................................................................................... 25
   3.2 Data in the Model ............................................................................................................ 25
   3.3 Benthic eDNA Data ......................................................................................................... 30
   3.4 Model Evaluation ............................................................................................................ 30
# TABLE OF CONTENTS

3.4.1 Predicted Versus Observed ................................................................. 30  
3.4.2 Model Validation .................................................................................. 30  
3.4.3 Sensitivity Analysis ............................................................................ 31  

4. RESULTS .................................................................................................................. 32  
4.1 Bayesian Network Relative Risk Method (BN-RRM) models .................... 32  
4.2 Water Quality Endpoints – Posterior Probability Distributions ............... 35  
4.3 Benthic Endpoints – Posterior Probability Distributions .......................... 37  
4.3.1 Comparison between Estuaries: Primary Producers ......................... 37  
4.3.2 Comparison between Sub Regions: Primary Producers ...................... 38  
4.3.3 Comparison between Estuaries: Primary Consumers and Decomposers 39  
4.3.4 Comparisons between Sub Regions: Primary Consumers and Decomposers 40  
4.4 Model Evaluation ......................................................................................... 41  
4.4.1 Predicted Versus Observed ................................................................. 41  
4.4.2 Model Validation .................................................................................. 43  
4.4.3 Sensitivity Analysis ............................................................................ 45  
4.5 Interactive Capability of Netica Models as Research Product .................... 49  

5. DISCUSSION ............................................................................................................. 50  
5.1 Water Quality and Benthic Endpoint Results .......................................... 50  
5.2 Uncertainty and Data Needs ....................................................................... 52  
5.3 Integrating eDNA into a Risk Assessment ................................................ 53  

REFERENCES ........................................................................................................... 54
# TABLE OF CONTENTS

SUPPLEMENTAL INFORMATION

BENTHIC ENVIRONMENTAL DNA ........................................................................................................... 59

S1. Methods - CSIRO ............................................................................................................................... 61

S1.1 CSIRO Field Sampling and eDNA Sequencing ............................................................................... 61

S2. Methods – Data Analysis for eDNA .................................................................................................. 63

S2.1 Big Picture Patterns in eDNA – Ordination .................................................................................. 63

S2.2 Incorporating eDNA into the BN-RRM Model as Endpoints ......................................................... 63

S2.2 Determining Water Quality Predictors for the eDNA Endpoints .................................................. 65

S3. Results ............................................................................................................................................... 67

S3.1 Big Picture eDNA Analysis – Ordination with NMDS ................................................................. 67

S3.2 Vector Fitting to NMDS Ordination ............................................................................................. 69

S3.3 BEST Test and Correlations Results ......................................................................................... 71

IMAGES OF THE NETICA BN-RRM MODELS ................................................................................... 73
LIST OF FIGURES

Figure 1  Relative Risk Model (RRM) framework………………………………………………………………………...3
Figure 2  Map of the South East Queensland (SEQ) region..................................................................................9
Figure 3  Map of the Noosa Estuary sub risk regions.........................................................................................14
Figure 4  Map of the Pine Estuary sub risk regions.........................................................................................15
Figure 5  Map of the Logan Estuary sub risk regions.........................................................................................17
Figure 6  Conceptual model structure.............................................................................................................24
Figure 7  Bayesian network relative risk model from Netica™ .................................................................34
Figure 8  Water quality endpoint posterior probability distributions............................................................35
Figure 9  Primary producer biotic endpoint posterior probability distributions.................................38
Figure 10 Primary consumer and decomposer posterior probability distributions..............................40
Figure 11 Comparison of BN results to observations for water quality endpoints.....................................42
Figure 12 Comparison of BN results to observations for biotic endpoints.....................................................43
Figure 13 Sensitivity analysis to the stressor variables..............................................................................47
Figure 14 Sensitivity analysis to the intermediate water quality variables...............................................48

LIST OF TABLES

Table 1  Environmental conditions of the South East Queensland (SEQ) estuaries.................................12
Table 2  SEQ risk assessment endpoints......................................................................................................20
Table 3  Methodology used to discretize model variables..........................................................................29
Table 4  BN model results for probability of achieving water quality objectives..................................36
Table 5  Cross-validation results for water quality and biotic endpoints....................................................44
LIST OF SUPPLEMENTAL INFORMATION

1. BENTHIC EUKARYOTE ENVIRONMENTAL DNA DATA ANALYSES

LIST OF FIGURES

Figure S1: Map of the SEQ region and five estuaries sampled for eDNA........................................60
Figure S2: NMDS plots of benthic MOTUs for 2010 and 2012............................................................68
Figure S3: NMDS plots of benthic MOTUs fit to environmental data for 2010 and 2012....................70

LIST OF TABLES

Table S1: EHMP water quality monitoring and CSIRO eDNA co-located sample sites..................62
Table S2: Average relative richness for benthic taxa groups..............................................................64
Table S3: eDNA benthic taxa group definitions..............................................................................65
Table S4: Summary of statistical analysis used to determine benthic taxa predictors....................72

2. IMAGES OF THE BN NETICA MODELS

Figure S4: Lower Logan BN..............................................................................................................73
Figure S5: Lower Noosa BN...........................................................................................................74
Figure S6: Middle Noosa BN.........................................................................................................75
Figure S7: Upper Noosa BN..........................................................................................................76
Figure S8: Lower Pine BN.............................................................................................................77
Figure S9: Middle Pine BN............................................................................................................78

The Bayesian network model files (.neta) are available electronically or upon request. Download the free version of Netica to view the models without a license (https://norsys.com/netica.html).
ACRONYMS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>Australian Bureau of Statistics</td>
</tr>
<tr>
<td>BN</td>
<td>Bayesian network</td>
</tr>
<tr>
<td>BN-RRM</td>
<td>Bayesian network-Relative risk method</td>
</tr>
<tr>
<td>Chl-a</td>
<td>Chlorophyll-a</td>
</tr>
<tr>
<td>CPT</td>
<td>Conditional Probability Table</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Commonwealth Scientific and Industrial Research Organization</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>eDNA</td>
<td>Environmental DNA</td>
</tr>
<tr>
<td>EHMP</td>
<td>Ecosystem Health Monitoring Program</td>
</tr>
<tr>
<td>EM-learning</td>
<td>Expectation-maximization learning</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic information systems</td>
</tr>
<tr>
<td>MOTU</td>
<td>Molecular Operational Taxonomic Unit</td>
</tr>
<tr>
<td>PPD</td>
<td>Posterior probability distribution</td>
</tr>
<tr>
<td>RRM</td>
<td>Relative risk method</td>
</tr>
<tr>
<td>SEQ</td>
<td>South East Queensland</td>
</tr>
<tr>
<td>SEQHWP</td>
<td>South East Queensland Healthy Waterways Partnership</td>
</tr>
<tr>
<td>STP</td>
<td>Sewage Treatment Plant</td>
</tr>
<tr>
<td>TN</td>
<td>Total nitrogen</td>
</tr>
<tr>
<td>TP</td>
<td>Total phosphorus</td>
</tr>
<tr>
<td>USEPA</td>
<td>United States Environmental Protection Agency</td>
</tr>
</tbody>
</table>
1. **INTRODUCTION**

In this research, I apply the Relative risk model with Bayesian networks (BN-RRM) to an integrated assessment of water quality and DNA-derived benthic communities for three estuaries in South East Queensland (SEQ), Australia. Risk to achieving regional water quality objectives was calculated using site-specific monitoring data to quantify relationships between salinity, climate, land use, water quality, and benthic communities. I used site specific data and 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 (DNA) data from sediments. The application of eDNA to risk assessment is largely unexplored; and to my knowledge this is the first attempt to synthesize DNA-derived measurements of biological composition into an ecological risk assessment framework.

1.1 **South East Queensland**

Human activities near coastal areas are changing the water quality and biota of the world’s estuaries. In Australia, more than 85% of the 22 million people live within 50km of the coast, and the Australian Bureau of Statistics (ABS 2003) predicts that the population will increase by 82% by 2056. Coincidently, intensive land use for development and associated stressors to water resources are also increasing. In SEQ, Australia’s fastest growing region, the combination of heavy rainfall and cleared land increases 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 some 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, monitoring and predicting the changes to estuarine water quality and biota provides valuable information for management of the region (Bunn et al. 2005; Moss et al. 2006).

Scientists can contribute to natural resource management by developing predictive models and assessments that link climate 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). Bayesian networks (BNs) are a modeling platform that is 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).

1.2 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 objectives. In a risk assessment, the management objectives are used to define the endpoints and ultimately drive the assessment (Landis and Wiegers 2005; Suter 2007). The size of the SEQ region warrants a framework that can incorporate complex ecosystem and multiple stressor interactions across habitats, space and time. The RRM has been used for nearly 20 years for landscape scale risk assessments to quantify the relative risk for multiple endpoints across sub regions of a site (Wiegers et al. 1998; Landis and Wiegers 2005; Ayre and Landis 2012; Hines and Landis 2014; Herring et al. 2015). Figure 1 illustrates the RRM framework that connects sources, stressors, habitats and impacts and emphasizes the importance of location (Landis and Wiegers 2005).
Figure 1. The Relative Risk Model (RRM) framework (Landis and Wiegers 2005) is used to organize cause-effect information and develop the Bayesian network (BN) model structure.

Most recently, the RRM has been used in conjunction with BNs because of the flexibility of the modeling platform, probabilistic nature of the calculations, and inherent representation of uncertainty (Ayre and Landis 2012). Hereafter, I refer to the combination of BNs and the RRM as a BN-RRM model. The BN-RRM has been used in a variety of ecological contexts including to examine risk of storm water runoff to Coho salmon in Puget Sound (Hines and Landis 2014), risk of 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. 2016, Johns et al. 2016).

1.3 Bayesian Network (BN) Modeling

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 predictive model (Varis et al. 1994; Borsuk et al. 2004).

1.4 Environmental DNA (eDNA)

For the present study, I integrated information from a variety of sources into a BN-RRM model to assess risk of intensive land use practices to water quality objectives and benthic communities in the SEQ estuaries. Of particular note is the use of 18S eukaryotic eDNA data for measuring the benthic community endpoint. 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; 2014; 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.

Recent advances in sequencing technology and bioinformatics provide an exciting opportunity to advance assessments 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 eDNA sampling provides insight into these responses, and can provide information for further understanding of taxa sensitivities to natural and anthropogenic stressors. In estuaries for example, 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 eDNA into context with management objectives and other information about the ecosystem is the motivation for this project.

1.5 Study Objectives

My objectives 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 patterns of 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.

I used the BN-RRM approach to quantify the risk of meeting water quality objectives and to predict benthic community structure for three estuaries in SEQ; the Noosa, the Pine, and the Logan. In more general terms, the BN-RRM model predicts patterns in water quality and organisms response to land use and climate.

1.6 Summary of Findings

Based on my study objectives, the three major findings of this work are:

1) I demonstrated the use of the BN-RRM approach to model 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 indicate 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 and photosynthesizing taxa richness, like Diatoms and Green Algae, 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 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.
2. SOUTH EAST QUEENSLAND (SEQ) STUDY REGION

Queensland is the second largest Australian state and covers the northeast quadrant of the continent. The SEQ region centers around the Queensland state capital of Brisbane and Moreton Bay (Figure 2). 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.

Moreton Bay is a large shallow embayment, separated from the ocean by sand islands, and it accumulates sediment, nutrients, and pollutants from the catchments than drain into it (Bunn et al. 2005; SEQHWP 2007a). The bay is a designated marine park and is listed as 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 has an average flushing rate (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 Estuary) (Dennison and Abal 1999; SEQHWP 2007a). Residence time is a measure of estuarine circulation and is an important factor in determining nutrients, turbidity, and dissolved oxygen levels in estuary waters (Ryan et al. 2003). Estuaries with shorter residence times generally have higher flushing rates of saline ocean waters, and thus lower nutrients and turbidity levels. The opposite is generally true for estuaries with long residence times.

Besides estuarine circulation, climate and land use are other important factors in determining water quality. An overview of the climate and land use in SEQ as they relate to water quality in estuaries is provided below. The BN-RRM model in this study focuses on two classes of
stressors: climate and land use, as two of the largest factors influencing the environmental condition in SEQ estuaries. In terms of management, land use is the major factor that can be managed.

2.1 Climate

The SEQ climate is subtropical with mild 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 indicate 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 20th 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 deposition of fine mud in Moreton Bay and the surrounding estuaries (O’Brien et al. 2012).
Figure 2. South East Queensland region with catchment boundaries and land use as mapped in 2012-2013. The Noosa, Pine, and Logan catchment boundaries are highlighted; the estuarine portion of these catchments were evaluated in this assessment.
2.1 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).

Overall, the largest land uses in the SEQ region are production from relatively natural environments, which includes grazing (56%) and conservation land (16%), but these uses predominate in upper to mid catchment areas. Intensive uses (shown in red in Figure 2) are concentrated near the coast and along the estuaries. Based on the Australian Land Use and Management Classification scheme, intensive land use is one of five primary land use classes and includes land uses that highly inhibit natural processes (State of Queensland 2010). Intensive land use is associated with complete or nearly complete removal of remnant vegetation. 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 class has increased by 9% as more land is being developed to meet the needs of a growing population (State of Queensland 2014).

Estuarine water quality is impacted by diffuse pollution from urban run-off, particularly during construction. The major components of diffuse source pollution across SEQ are sediment, nitrogen and phosphorus (Cottingham et al. 2010). The estimated loadings of pollutants per unit area of urban land use are 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, I am also using percent intensive land use for this study.
2.2 Management of SEQ Waterways

Management of the SEQ waterways is a joint effort between the State of Queensland, local governments, and the South East Queensland Healthy Waterways Partnership (SEQHWP), a non-profit 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.

Since 1999, the EHMP has routinely collected water quality and biologic monitoring data from freshwater, estuarine and marine sites in SEQ. Using an index calculated with the EHMP data, the Healthy Waterways Partnership produces an annual report card for the freshwater and estuary 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 grade calculated for the estuary portions of the catchment does not take into account any biological data (e.g. benthic organisms or fish) due to the costs and turnaround time in sample processing. Environmental DNA may provide a rapid and cost effective way to gather ecological data for SEQ estuaries.

The three SEQ estuaries evaluated in this risk assessment, the Noosa, Pine and Logan (Table 1), represent a range of land use impact and geomorphologies in the SEQ. I 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 within the estuaries, were delineated by the State of Queensland and made available via GIS shapefiles. My 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. Summary of environmental conditions of the three South East Queensland estuaries used in this assessment and the 2015 Healthy Waterways Partnership report cards.

<table>
<thead>
<tr>
<th>Estuary</th>
<th>Total Catchment Area (km²)</th>
<th>Estuary Catchment Area (km²)</th>
<th>Estuary Type⁠¹</th>
<th>Major Energy Influence¹</th>
<th>2015 EHMP Grade²</th>
<th>Land use and Sources of Anthropogenic Stressors²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noosa</td>
<td>854</td>
<td>251</td>
<td>Wave-dominated Estuary</td>
<td>Wave Energy</td>
<td>A-</td>
<td>Urban areas concentrated around lower estuarine reaches and occupy less than 3% of total catchment; No point sources.</td>
</tr>
<tr>
<td>Pine</td>
<td>825</td>
<td>67</td>
<td>Tide-dominated Estuary</td>
<td>Tide energy</td>
<td>C</td>
<td>Estuarine reaches are highly urbanized; 2 impoundments forming lakes on North Pine River; WWTP.</td>
</tr>
<tr>
<td>Logan</td>
<td>3,076</td>
<td>302</td>
<td>Tide-dominated Delta</td>
<td>Tide energy</td>
<td>D</td>
<td>Estuarine reaches flow through rural residential and urban areas; several aquaculture facilities are located along the banks of the Logan near its mouth; two WWTP.</td>
</tr>
</tbody>
</table>

Notes:
1. Estuary Type and Major Energy Influence was designated by Geoscience Australia (Ryan et al. 2003).

2.3 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 (Figures 2 and 3). 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). The Noosa is classified as a wave-dominated estuary (Table 1), and as such, the Lower Noosa risk region is dominated by wave energy, low turbidity, and sandy
sediments (Figure 3). The Middle and Upper regions of the Noosa are dominated by freshwater flows and lower salinities than in a Tide-dominated estuary like the Pine or Logan. There are no published residence times for the Noosa Estuary; but given that the Noosa is connected directly into the Pacific Ocean, the residence times are likely shorter than for the Pine and Logan, which both discharge into the bay.

There are no point sources discharging to estuarine waters in the Noosa, or dams in the upper reaches (Figure 3). The Healthy Waterways Partnership regularly gives the Noosa the highest grades (Table 1) and considers the overall environmental condition of the Noosa excellent (SEQHWP 2015). The majority of the intensive land uses are located in the Lower Noosa region, where many hotels, vacation homes and businesses are located. Upstream, the Middle and Upper regions consist largely of conservation land with intact mangrove forests lining the waterways.
2.4 Pine

While the Pine catchment is similar in overall area to the Noosa, the Pine Estuary area is smaller than the Noosa Estuary (Table 1). The Pine catchment is located to the north of Brisbane and includes two tributaries, the North and South Pine Rivers (Figure 4). Both rivers originate from undeveloped regions of protected forestland 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 Pine is classified as a tide-dominated estuary and has a residence time of 55-62 days (Dennison and Abal 1999; Ryan et al. 2003). Tide-dominated estuaries are generally characterized by high turbidity due to turbulence induced by the tides.

The Murrumba Downs Sewage Treatment Plant (STP) discharges into the North Pine River approximately 10 km from the estuary mouth (Figure 4) and has not been upgraded since 2000 when EHMP monitoring began. The estuary flows into Hayes Inlet that contains mangrove and seagrass habitat important to recreational and commercial fisheries, as well as migrating birds. The Healthy Waterways Partnership gave the Pine a C grade in 2015, and considers the environmental conditions to be fair (SEQHWP 2015).

Figure 4: Pine Estuary sub risk regions: Lower Pine and Middle Pine.
2.5 Logan

The Logan is located just south of the city of Brisbane is the second largest catchment in the SEQ (Figure 5). The mouth of the Logan Estuary lies at the southern end of the Moreton Bay, and the estuary has a relatively long residence time of 66-75 days (Dennison and Abal 1999). Similar to the Pine Estuary, the Logan is classified as tide-dominated and is characterized by high turbidity and strong tidal currents. The mouth of the Logan Estuary is more filled in with sediments than the Pine, and a delta with sand bars and channels has formed at the mouth (Ryan et al. 2003).

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 (Figure 5).

The Healthy Waterways Partnership gave the entire Logan catchment a D grade in 2015, and considers the environmental conditions to be poor (SEQHWP 2015).
Figure 5: Logan Estuary sub risk regions: Lower Logan and Middle Logan.
3. METHODS

This section first provides an overview 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 in the Supplemental Information.

3.1 Relative Risk Model (RRM) Overview

The RRM methodology described in Landis and Wiegers (2005) was used for selecting risk assessment endpoints and developing the conceptual model (Figure 1). Additional methods outlined in Ayre and Landis (2012) describe the use of BNs within the RRM framework for modeling and risk calculations. I have bolded key terms relative to BN-RRM process.

The first step in the RRM is to solicit ecological values from stakeholders and collaboratively decide on endpoints for the risk assessment. Ideally, the selected endpoints have management goals associated with them so the risk assessor can develop the model to be specific. Next, all plausible stressors to the endpoints and the sources of those stressors are identified through further solicitation and research (Landis and Wiegers 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 Wiegers 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 Wiegers 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 Wiegers 2005). All variables are defined with an appropriate measurement attribute (e.g. % DO saturation or mm of rainfall per 30-
day period). The variables are then discretized into ranked states (this step further described in Section 3.2), and risk is calculated as a probability of achieving a given risk state. Evaluation of the model includes a sensitivity analysis and comparison to observed conditions, if that information is available. Finally, model results are communicated in a fashion that portrays the relative risk to endpoints in the context of management goals (Landis and Wiegers 2005).

3.1.1 Selection of Endpoints

During the conceptual model development, risk assessors and modelers should host stakeholder meetings to clarify management goals, select ecological endpoints and solicit feedback (Borsuk et al. 2004; Landis and Wiegers 2005). However, given the information currently available about stakeholder values and limitations in resources; I relied on a less formal process to determine endpoints. This process 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 Queensland government research efforts, and the results of those surveys were publically available (Abal et al. 2005; SEQHWP 2007a; SEQHWP 2007b; Healthy Waters 2013). Based on the information available, I 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 model that I 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 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 entities (USEPA 1998). In SEQ, the regional Water Quality Objectives are the entity and DO and Chl-a are the attributes (Table 2). 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 (Table 2).

Table 2. Entity and attributes of the SEQ risk assessment endpoints.

<table>
<thead>
<tr>
<th>Entity</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Quality Objectives</td>
<td>Dissolved oxygen (% saturation)</td>
</tr>
<tr>
<td></td>
<td>Chlorophyll-a (µg/L)</td>
</tr>
<tr>
<td>Benthic Community Assemblage (no regional management objectives available)</td>
<td>Diatom relative richness</td>
</tr>
<tr>
<td></td>
<td>Dinoflagellates relative richness</td>
</tr>
<tr>
<td></td>
<td>Fungi relative richness</td>
</tr>
<tr>
<td></td>
<td>Meiofauna relative richness</td>
</tr>
<tr>
<td></td>
<td>Protist – Protozoan relative richness</td>
</tr>
<tr>
<td></td>
<td>Green Algae relative richness</td>
</tr>
</tbody>
</table>

Dissolved oxygen and Chl-a were selected for this risk assessment because they are the two most common water quality responses to increased nutrient loading, can indicate eutrophication, and can negatively affect other estuarine biota and human health (Cloern 2001). Increases in phytoplankton biomass, 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 respiration. When phytoplankton blooms are too intense, they can 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 (Table 2). Benthic communities are commonly monitored in waterways, and many studies demonstrate their response 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). Measures of biodiversity, such as richness, are commonly used to assess the condition of benthic communities. In estuarine macrobenthic studies, eutrophication (and human disturbance in general) tends to lead to a pronounced reduction in richness (Johnston and Roberts 2009). However in DNA-derived estuarine benthic studies that identify a much broader taxonomic scope of organisms, this trend is not always the case. In fact, the most highly disturbed estuaries sometimes have the highest richness due to inputs from the adjacent catchment (Chariton et al. 2015). Thus, we can expect to see differences in DNA-derived studies versus traditional benthic studies, and there is much work to be done to understand the community response in eDNA studies.

There are currently no regional objectives to assess risk of achieving objectives for the eukaryote benthic community endpoints, so instead the endpoints reflect the patterns of benthic assemblages as a result of stressors to each estuary. The eukaryotic benthic dataset includes over 8,000 unique molecular operational taxonomic units (MOTU) sequence reads. All of the taxa associated with the MOTUs 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. Richness was calculated as the count of unique MOTU reads per sample (See Supplemental information for more discussion).

### 3.1.2 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 reason the 2012 data did not detect contaminants was likely due to January 2011 flooding (see Section 2.1), which flushed sediments from the estuaries into Moreton Bay (van den Honert and McAneney 2011). Excess sediment and nutrient in waterways were retained as stressors for the BN-RRM.

Currently, the major anthropogenic source of nutrient and sediment loading to waterways is diffuse runoff from intensive land uses (Cottingham et al. 2010). These intensive land uses 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 such as poorly functioning sewage treatment plants and aquaculture discharges have largely 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). Other major contributing factors to nutrients and suspended sediments in SEQ estuarine waters is the estuarine circulation from the Pacific Ocean (measured by salinity), and heavy summer rainfalls that create the conditions for diffuse runoff
events. These environmental sources (salinity, rainfall, and season) plus the anthropogenic source (intensive land use) are used as the four inputs to the BN-RRM model. The inputs for each risk region were based on site-specific data.

### 3.1.3 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, and downloaded as GIS shapefiles. I 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 set different water quality objectives for the lower (more saline) versus middle (fresher water) sections of the estuaries, so I delineated those regions boundaries 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 3-5 present the different risk regions of the estuaries, point sources, and EHMP and eDNA sample locations. As the figures show, the eDNA and EHMP locations are co-located in most instances.

### 3.1.4 Model Structure

The structure of the BN-RRM model for this study (Figure 6) was developed based on the original RRM framework (Landis and Wiegers 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. 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 and State of Queensland scientists.

![Conceptual model structure](image)

**Figure 6.** Conceptual model structure; the network structure is the same for all risk regions.

Stressors and sources of those stressors (Season, Intensive Land use, Total Monthly Rainfall, and Surface Water Salinity) are the input nodes and start the pathway on the left side of the model. The input nodes predict the intermediate water quality nodes (Total Nitrogen [TN], Total Phosphorous [TP], Turbidity, and Temperature). The intermediate nutrients, turbidity and temperature nodes describe the effect of the stressors on the water quality habitat. The
intermediate nodes link to the endpoint nodes, which describe the predicted impact to the attributes of the valued ecological entity (Table 2).

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 communities 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.

3.1.5 Model Assumptions

All models have exactly the same physical BN structure (Figure 6), meaning that I assume cause-effect relationships between variables in the SEQ estuaries are generally the same. This works because the models were built based on a breadth 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.

3.2 Data in the Model

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.

1) Discretizing the nodes.

Each node in the model was discretized into states or ranges (Table 3). The goal of discretizing was to represent the influences of the variables on the endpoint with the fewest discrete states necessary. In most cases, I created four states corresponding to the zero, low, med, high risk ranking scheme described by Landis and Wiegers (2005), and used by many other BN-RRMs (Ayre and Landis
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, states were established using classifications recognized by estuary scientists or local management objectives. I have included the discretized ranges and justifications for each variable in Table 3. For water quality variables that have a management objective (TN, TP, Turbidity, Chl-a, DO), the lowest possible risk state(s) 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, which is usually from 85 - 105% saturation (Table 3).

The discretization of all input, intermediate, and benthic community nodes is exactly the same for all models and estuaries. The two water quality objectives endpoint nodes vary slightly between regions based on differences in regional management objectives. Definitions of the variables, sources of the data, and justifications for the breakpoints are provided in Table 3. I 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.

2) Parameterizing 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 stored within child nodes and are represented as a matrix of probabilities of a child node state occurring given the state of its parents (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, 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 only one method: case files and the expectation-maximization learning (EM-learning) algorithm, and an automated function available in Netica (Norsys 2014). I 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 synoptic findings that go together to provide information about the variables (nodes) in the model. For the BN-RRM models, the case files were composed of land use and monitoring data and each case included all of the measurements for a given sample. The EM-learning algorithm iteratively calculates the maximum likelihood estimates for the nodes in the model given the case file data and the model structure (Figure 6). Expectation-maximum learning has been routinely used for other environmental BN modeling (Pollino et al. 2007; Lucena-Moya et al. 2015), and was selected over other algorithms because it deals well with missing data.

I used site-specific data to create a unique case file for each estuary and parameterize the CPTs. Because there was so much monitoring data available for SEQ estuaries, I was able to use a unique case file for each estuary and the relationships in the models are estuary specific. The three case files (one for each estuary) consisted of EHMP monitoring data that had been collected on a monthly basis from 1999-2014 at regular locations within the estuary (Figures 3-5). Each sample (or case) in the case file was matched to a corresponding 30-day rainfall total (mm) and percent 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 relationships between all nodes except for the benthic richness endpoints.

The benthic nodes of the model were parameterized separately from other nodes using the eDNA data matched to corresponding EHMP monitoring data. For the benthic taxa, a single case-file of richness and water quality predictors from all five estuaries sampling for eDNA in SEQ (n = 287 cases) was used to parameterize the CPTs for all estuaries. This means the benthic CPTs in each
estuary and risk region are exactly the same. 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.

3) Setting the input (stressor) distributions.

The input nodes (Season, Rainfall, Land use and Salinity) for each of the seven risk regions were set with region-specific input distributions. The Rainfall and Season stressor nodes are the same for the sub regions of an estuary, but the Salinity and Land Use nodes differ among risk regions. The distributions for the Season and Salinity nodes were set based on EHMP monitoring data collected from 1999-2014. The distribution for the Intensive Land Use node was set based on 2012-2013 Queensland land use data. Finally, the Rainfall node distribution was set based on observations by the Queensland Bureau of Meteorology from 1985 - 2015 for the nearest weather station to the estuary.
Table 3. Methodology used to discretize model variables and the states of these variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discretization Methodology and Justification</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stressor Nodes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land Use (% Intensive Uses)</td>
<td>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.</td>
<td>&lt; 30 %</td>
</tr>
<tr>
<td>Total Monthly Rainfall</td>
<td>States were determined using natural breaks 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).</td>
<td>0 - 50 mm</td>
</tr>
<tr>
<td>Season</td>
<td>Austral Seasons</td>
<td>Autumn (Sept-Nov)</td>
</tr>
<tr>
<td>Salinity</td>
<td>Venice System for Classification of Marine Waters (Venice system, 1958).</td>
<td></td>
</tr>
<tr>
<td><strong>Intermediate Nodes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Nitrogen Concentration</td>
<td>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 high in a survey in Bricker et al. (2003). The middle states were discretized based on natural breaks of EHMP monitoring data.</td>
<td>&lt; 0.3 mg/L (Objective)</td>
</tr>
<tr>
<td>Total Phosphorous Concentration</td>
<td>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).</td>
<td>&lt; 0.03 mg/L (Objective)</td>
</tr>
<tr>
<td>Water Temperature</td>
<td>The states were discretized based on natural breaks of the EHMP monitoring data (Jenk’s).</td>
<td>13 - 19 °C</td>
</tr>
<tr>
<td>Turbidity</td>
<td>The lowest state (&lt; 8 NTU) meets default trigger values for SEQ (ANZECC/ARMCANZ 2000). The highest state (&gt;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.</td>
<td>&lt; 8 NTU (Objective)</td>
</tr>
<tr>
<td><strong>Endpoint Nodes - Water Quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissolved Oxygen Saturation</td>
<td>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 (&lt;50% sat.) is considered hypoxic (Breitburg 2002). The remaining states were set based on natural breaks (Jenk’s) that were rounded.</td>
<td>&lt; 50 % sat.</td>
</tr>
<tr>
<td>Chlorophyll-a Concentration</td>
<td>Based on region-specific objectives, the Objective states varies between risk region in the Noosa and Pine estuaries. See Table 4 for all regional Objectives for the endpoints. The states &gt;4 µg/L were set based on low, medium, high and hypereutrophic ranges reported in a survey in Bricker et al. 2003.</td>
<td>&lt; 2 µg/L (Objective)</td>
</tr>
</tbody>
</table>

**Endpoint Nodes - Benthic Relative Richness**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discretization Methodology and Justification</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diatom, Meiofauna, and Protozoan</td>
<td>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.</td>
<td>0 - 5 %</td>
</tr>
<tr>
<td>Dinoflagellate, Green Algae and Fungi</td>
<td>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.</td>
<td>0 - 2.5 %</td>
</tr>
</tbody>
</table>
3.3 Benthic eDNA Data

Scientists from CSIRO, Australia’s National Science Agency, collected the samples and sequenced the benthic eDNA 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 andCurrumin (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. A summary CSIRO’s methods and my analyses used to determine predictors is provided in the Supplemental Information.

3.4 Model Evaluation

I 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).

3.4.1 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, I compared the modeled BN mean value to the observed average value from data for the water quality and benthic endpoints.

3.4.2 Model 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 cases. For each estuary, I used 80% of the case file data to train the model and the remaining 20% to test the model (Pollino et al. 2007;
Chen and Pollino 2012). Netica processes the test cases one by one and compares the state the model predicts to the observed state from the case file. The error rates for the training cases were compared across endpoints and between estuaries.

3.4.3 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 important in predicting 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 model entropy (Marcot 2012).
4. RESULTS

4.1 Bayesian Network Relative Risk Method (BN-RRM) Models

The BN-RRM model for each estuary has two layers of predictions (Figure 7). The first layer uses climate stressors (Season and Monthly Rainfall), 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 (Figure 1). Using Netica (Norsys 2014), I created a separate BN model for each risk region that includes region specific relationships between the variables and unique stressor distributions.

The BN model results for the endpoints are summarized in three ways. First there are the predicted probabilities of a given endpoint states occurring (Table 3). These state 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 I have summarized the PPDs for all endpoint nodes (Figures 8, 9, and 10). 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 (≥ 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, I have compared the expected value to the observed mean value for each risk region as part of the model evaluation process (Figures 11 and 12). 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.
Figure 7. Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Logan Estuary with the Middle Logan risk region stressors selected.
4.2 Water Quality Endpoints – Posterior Probability Distributions

For both the DO and Chl-a endpoints, the percent probability of achieving water quality objectives is higher in the Noosa Estuary than in the Pine and Logan Estuaries (Table 4). Within the Noosa Estuary, the Lower, Middle and Upper sub regions have very similar PPDs and high certainty of achieving the objective risk state(s) (between 73-91% probability of achieving the Objective risk states for both DO and Chl-a) (Table 4). 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) (Figure 8). General trends included that 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 4). 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 8).

Figure 8. Water quality endpoint PPDs from the BN-RRM for each region and estuary. The dashed lines indicates the regional water quality objective range. The objectives differ slightly between regions (see Table 4 for exact numerical objectives).
Table 4. Bayesian network model results for the probability of achieving regional water quality objectives for each risk region. The objectives for each region are also shown.

### a) Dissolved Oxygen Endpoint

<table>
<thead>
<tr>
<th>Risk Regions</th>
<th>Objective</th>
<th>Probability to achieve objective</th>
<th>Relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noosa Lower (NL)</td>
<td>90 - 105%</td>
<td>74%</td>
<td>Medium</td>
</tr>
<tr>
<td>Noosa Middle (NM)</td>
<td>85 - 105%</td>
<td>81%</td>
<td>Low</td>
</tr>
<tr>
<td>Noosa Upper (NU)</td>
<td>85 - 105%</td>
<td>75%</td>
<td>Low</td>
</tr>
<tr>
<td>Pine Lower (PL)</td>
<td>90 - 105%</td>
<td>55%</td>
<td>Medium</td>
</tr>
<tr>
<td>Pine Middle (PM)</td>
<td>85 - 105%</td>
<td>24%</td>
<td>High</td>
</tr>
<tr>
<td>Logan Lower (LL)</td>
<td>85 - 105%</td>
<td>69%</td>
<td>Medium</td>
</tr>
<tr>
<td>Logan Middle (LM)</td>
<td>85 - 105%</td>
<td>16%</td>
<td>High</td>
</tr>
</tbody>
</table>

### b) Chlorophyll-a Endpoint

<table>
<thead>
<tr>
<th>Risk Regions</th>
<th>Objective</th>
<th>Probability to achieve objective</th>
<th>Relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noosa Lower (NL)</td>
<td>&lt; 1.8 µg/L</td>
<td>73%</td>
<td>Medium</td>
</tr>
<tr>
<td>Noosa Middle (NM)</td>
<td>&lt; 2.2 µg/L</td>
<td>77%</td>
<td>Low</td>
</tr>
<tr>
<td>Noosa Upper (NU)</td>
<td>&lt; 5 µg/L</td>
<td>91%</td>
<td>Low</td>
</tr>
<tr>
<td>Pine Lower (PL)</td>
<td>&lt; 2 µg/L</td>
<td>26%</td>
<td>Medium</td>
</tr>
<tr>
<td>Pine Middle (PM)</td>
<td>&lt; 4 µg/L</td>
<td>43%</td>
<td>High</td>
</tr>
<tr>
<td>Logan Lower (LL)</td>
<td>&lt; 4 µg/L</td>
<td>60%</td>
<td>Medium</td>
</tr>
<tr>
<td>Logan Middle (LM)</td>
<td>&lt; 4 µg/L</td>
<td>52%</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Note:

In the Lower Logan risk region, the probability of achieving water quality objectives for DO was surprisingly high (Table 4; 69% probability), given that the Logan Estuary is regularly given the lowest report card grades (given a D in 2015) by the SEQ Healthy Waterways Partnership. The Lower Logan may have higher DO because it has a very wide central basin at the mouth of the estuary, which contributes to greater renewal of DO in the surface waters via wind and tidal mixing (Ryan et al. 2003; Nezlin et al. 2009).
4.3 Benthic Endpoints – Posterior Probability Distributions

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

- Primary Producers (photosynthetic): Diatoms, Dinoflagellates, 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 natural and anthropogenic inputs (land use, climate, and salinity) to the estuaries. The patterns between risk regions reflect changes along the salinity gradient with the Lower risk regions representing marine regions and the Middle/Upper region representing less saline regions. Because there are not specific management objectives for richness of benthic eukaryote taxa in SEQ estuaries, there are no probabilities to achieve objectives to present or relative risk estimates.

4.3.1 Comparison between Estuaries: Primary Producers

The primary producer taxa groups have very different PPDs between estuaries, with the Diatoms and Green Algae having the largest differences (Figure 9). In the Noosa, 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.

In the Pine and Logan Estuaries, Diatom richness is lower, and Green Algae and Dinoflagellate richness is generally higher compared to the Noosa Estuary (Figure 9). 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 (Figure 9).

![Graphs showing relative richness and probability of state for different regions and taxonomic groups.](image)

**Figure 9.** Primary producer benthic endpoint posterior probability distributions for each risk region.

### 4.3.2 Comparison between Sub Regions: Primary Producers

Similar to the water quality endpoints, 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 9). The Logan and Pine PPDs, however, are different between risk regions, with Diatom richness decreasing from marine to fresh waters, while Green Algae and Dinoflagellates increases from marine to fresher waters (Figure 9).
4.3.3  *Comparison between Estuaries: Primary Consumers and Decomposers*

For the non-photosynthesizing taxa groups, only the Fungi taxa have markedly different patterns between estuaries. In the Noosa, relative richness is predicted to be low (in the 0 - 2.5% richness state) and that prediction is fairly certain (≥60% probability of the low state occurring) (Figure 9). In the Pine and Logan Estuaries, the PPDs indicate 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 9).

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 15), reflecting uncertainty about which richness state may occur and also reflecting the high variability of Meiofauna and Protozoan richness between samples in the estuaries (Figure 10).
4.3.4 **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) (Figure 14). 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.
result is similar to patterns observed for the water quality endpoints and photosynthetic benthic
taxa groups, and is a major finding in this work.

4.4 Model Evaluation

4.4.1 Predicted Versus Observed

Observed water quality data for each risk region was plotted against expected mean predictions
from the BN-RRM model (Figure 11). 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 11). 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
3), is likely skewing the expected mean high.

Observed benthic richness data for each risk region were plotted against the expected mean
richness predictions from the BN-RRM (Figure 12). The BN model predictions match the observed
richness trends more accurately for the photosynthetic groups (Diatoms, Dinoflagellates, and Green
Algae) than the non-photosynthetic groups (Protozoan, Meiofauna, and Fungi; Figure 12). Of the
non-photosynthetic groups, the BN model predicts Fungi richness trends the best (Figure 12).
Figure 11. 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.
Figure 12: 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.

4.4.2 Model Validation

Cross validation error rates for the water quality endpoints ranged from 33 - 60% (Table 5). Error rates were lowest in the Noosa Estuary models compared to the Pine and Logan models and were lower for the DO endpoint compared to Chl-a endpoint (Table 5). The Chl-a error rates are high for the Pine and Logan (56 and 59% error respectively) largely because the model fails to predict the difference between <2 µg/L and 2-4 µg/L states. The Pine and Logan have consistently higher Chl-a values than the Noosa, and the <2 µg/L state is not that relevant to management because Chl-a
objectives are <4 µg/L with the exception the Lower Pine (see Table 4 for all regional objectives).

The Noosa, on the other hand, has consistently very low Chl-a concentrations, so the model predicts those low Chl-a states better and that is reflected in the lower error rates for the Noosa (32%; Table 5).

Error rates for the benthic endpoints were comparable to the water quality endpoints and ranged between 28 and 40% (Table 5). All of the estuaries have the same error rates for the benthic endpoint because the benthic CPTs were parameterized with the same case file (See Section 3.2 for more explanation of why this is.). Lucena-Moya et al. (2015) reported error rates ranging from 20 to 52% for a similar study using case-learning to parameterize the CPTs for a model that predict macrobenthic richness. The high average error rates across all models and endpoints (>30%) likely reflects the large amount of variability within both the water quality and benthic eDNA data.

**Table 5.** Validation results for the water quality and benthic taxa endpoints are shown.

<table>
<thead>
<tr>
<th>Water Quality Endpoints</th>
<th>Cross-validated error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Noosa</td>
</tr>
<tr>
<td>Dissolved oxygen</td>
<td>32.47</td>
</tr>
<tr>
<td>Chlorophyll-a</td>
<td>32.66</td>
</tr>
<tr>
<td>Mean error rate across all models (standard error)</td>
<td>32.6 (0.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benthic Taxa Endpoints</th>
<th>All estuaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diatom</td>
<td>28.07</td>
</tr>
<tr>
<td>Dinoflagellates</td>
<td>40.35</td>
</tr>
<tr>
<td>Green Algae</td>
<td>40.35</td>
</tr>
<tr>
<td>Meiofauna</td>
<td>35.09</td>
</tr>
<tr>
<td>Fungi</td>
<td>29.82</td>
</tr>
<tr>
<td>Protozoan</td>
<td>33.33</td>
</tr>
<tr>
<td>Mean error rate across all models (standard error)</td>
<td>34.5 (2.1)</td>
</tr>
</tbody>
</table>
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 percent 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 presented based on the final model configuration. I divided the sensitivity results into two categories: sensitivity to inputs/stressors (Figure 13), and sensitivity to intermediate water quality variables (Figure 14).

**Sensitivity to Inputs/Stressors.** The Salinity stressor was clearly the most influential value for all endpoints except for Chl-a, where Season was more influential (Figure 13). Sensitivities to the Land Use and Rainfall stressors varied widely between the endpoints, but in general had lower mutual info compared to Salinity (Figure 17). Outcomes of the sensitivity analysis concur with estuarine science that the horizontal salinity gradient and estuarine circulation is a major factor in determining nutrients and DO levels as well as biotic assemblages (Ryan et al. 2003).

All of the Noosa BN-RRM model endpoints tended to be more sensitive (high percent mutual information) to Salinity than the other estuaries (Figure 13). This result is consistent with the results of Chariton et al. (2015), and multivariate analysis that I performed of the eukaryotic MOTUs and environmental variables (see Supplemental Information). In those analyses, the Noosa estuary eukaryotic benthos tended to be strongly associated with the salinity gradient.

**Sensitivity to Intermediate Variables.** The most influential water intermediate water quality variables for all endpoints were the nutrients TN and TP, as these two intermediate water quality variables linked to each endpoint. Total nitrogen had slightly higher mutual information than TP for all
endpoints (Figure 14). 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).
Figure 13. Mutual info percent from sensitivity analysis for the endpoints to the four input variables. Results are shown for all regions. Please note different x-axis scale for each endpoint.
Figure 14. Mutual info percent from sensitivity analysis for the endpoints to the four intermediate water quality variables. Results are shown for all regions. Please note different x-axis scale for each endpoint.
4.5 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. An image of the Logan model is presented in Figure 7, and the Pine and Noosa models are presented in the Supplemental Information. Netica files (.neta) will be also be provided 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 one can select Intensive Land use to all be in the 65 - 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 is lower DO, higher Chl-a, lower Diatoms, and higher Fungi.

There are limitations to which stressor scenarios you can test, however. It currently 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 - 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 and with other sources of information including primary literature, other model predictions, or expert elicitation.
5. 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 are simpler than other estuary models, but they accurately predict the current patterns of water quality and relative richness 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 models 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.

5.1 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 to water quality 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 (Figures 9 and 10). 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 50% probability of achieving water quality objectives. These regions include the Noosa and Lower Pine and Logan regions. Trends among the risk regions include:

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

Thus, these biotic signals indicate 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 are major decomposers of woody and herbaceous substrates in marine and estuarine ecosystems. Higher levels of Intense land use in the Pine and Logan Estuary catchments as well as the sewage treatments plants are the cause of the higher nutrients, and consequently more organic matter entering the waterways. Thus, Fungi richness might be 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 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 has been mentioned previously; there are no 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 to 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.

5.2 Uncertainty and Data Needs

All knowledge is uncertain, and BNs explicitly express uncertainty by representing all variables as probability 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 two factors: 1) data were collected from only two sampling events and 2) there is high natural variability of benthic eukaryote communities. 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 benthic organisms are populations that can fluctuate due to many reasons including changes in water quality and labile food inputs, more years and seasons of data would further inform the model.

On the other hand, there are more than 15 years of sampling data for water quality or rainfall in SEQ, 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, I am 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 those relationship are further refined, managers could quickly quantify how much management would have to occur to achieve desirable endpoint results.

5.3 Integrating eDNA into a Risk Assessment

While there is still much work to be done to fully capture the benthic community eDNA data, 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 learning algorithms and case-files, the relationships between the eDNA and predictor variables can be quickly determined based on site specific monitoring data.

Bayesian network modeling is meant to be an iterative process where information can be added as it becomes known or new data are collected. As new EHMP data are 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.
REFERENCES


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).


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 supplemental information 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, and Turbidity), Salinity, and benthic community richness.
Figure S1. Overview map of the five estuaries (Noosa, Maroochydore, Pine, Logan, and Currumbin) and the eDNA benthic sample locations from the CSIRO 2010 and 2012 field collection events.
S1. METHODS - CSIRO

S1.1 CSIRO Field Sampling and eDNA Sequencing

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 S1). 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 main text Figures 3-5).
Table S1: Lookup table for the co-located EHMP water quality sites and eDNA sample sites in the Logan, Pine, and Noosa Estuary risk regions.

<table>
<thead>
<tr>
<th>Risk Region</th>
<th>EHMP estuary sites</th>
<th>2010 eDNA sites</th>
<th>2012 eDNA sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Logan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>LL1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>LL2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>211</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>212</td>
<td>L1</td>
<td>LL3</td>
<td></td>
</tr>
<tr>
<td>Middle Logan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>202</td>
<td>L2</td>
<td>LL4</td>
<td></td>
</tr>
<tr>
<td>203</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>204</td>
<td>L3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>205</td>
<td>L4</td>
<td>LL5</td>
<td></td>
</tr>
<tr>
<td>206</td>
<td>L5</td>
<td>LL6</td>
<td></td>
</tr>
<tr>
<td>207</td>
<td></td>
<td>LL7</td>
<td></td>
</tr>
<tr>
<td>Lower Pine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>PP1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>801</td>
<td>P1</td>
<td>PP2</td>
<td></td>
</tr>
<tr>
<td>802</td>
<td>P2</td>
<td>PP3</td>
<td></td>
</tr>
<tr>
<td>Middle Pine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>803</td>
<td>P4</td>
<td>PP5</td>
<td></td>
</tr>
<tr>
<td>804</td>
<td>P5</td>
<td>PP6</td>
<td></td>
</tr>
<tr>
<td>811</td>
<td>P3</td>
<td>PP4</td>
<td></td>
</tr>
<tr>
<td>812</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>814</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower Noosa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1601</td>
<td>NN1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1603</td>
<td></td>
<td>NN2</td>
<td></td>
</tr>
<tr>
<td>Middle Noosa</td>
<td>1604</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1611</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1613</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1614</td>
<td>N2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1616</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1617</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1624</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1626</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1636</td>
<td>N1</td>
<td>NN3</td>
</tr>
<tr>
<td>Upper Noosa</td>
<td>1608</td>
<td>N3</td>
<td>NN5</td>
</tr>
<tr>
<td></td>
<td>1609</td>
<td>N5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1610</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1615</td>
<td>N4</td>
<td>NN6</td>
</tr>
<tr>
<td></td>
<td>1618</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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-CATCTAAGGGGCATCACAGACC (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/) (Chariton et al. 2015).

S2. METHODS – DATA ANALYSIS FOR EDNA

S2.1 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, I 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.

S2.2 Incorporating eDNA into the BN-RRM Model as Endpoints

To incorporate eDNA information in the BN model, I normalized the 2010 and 2012 samples into a single dataset by classifying the MOTUs into 14 taxa groups based on their phylum classification
(Table S2). Then I calculated the relative richness of each group by summing the MOTU presence/absence data for each sample per group, and dividing by the total MOTUs per sample. The taxa group with the highest relative richness (34%) for both the 2010 and 2012 samples was the Unclassified organisms, followed by the Protozoans, Diatoms and Meiofauna. For the benthic endpoints, I selected the six taxa groups with the highest MOTU richness in both 2010 and 2012, excluding the Unclassified group (Table S2). The six groups were sorted based on taxonomic information from the SILVA database, and the associated Kingdom, Phylum, and other phylogenetic information is provided in Table S3.

**Table S2:** 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.

<table>
<thead>
<tr>
<th>2010 (in order of richness)</th>
<th>Average relative richness</th>
<th>2012 (in order of richness)</th>
<th>Average relative richness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Unclassified</td>
<td>34</td>
<td>1  Unclassified</td>
<td>34</td>
</tr>
<tr>
<td>2  Protist - Protozoan*</td>
<td>16</td>
<td>2  Protist – Protozoan*</td>
<td>15</td>
</tr>
<tr>
<td>3  Diatom*</td>
<td>14</td>
<td>3  Diatom*</td>
<td>15</td>
</tr>
<tr>
<td>4  Meiofauna*</td>
<td>10</td>
<td>4  Meiofauna*</td>
<td>10</td>
</tr>
<tr>
<td>5  Green Algae*</td>
<td>5.6</td>
<td>5  Dinoflagellates*</td>
<td>5.2</td>
</tr>
<tr>
<td>6  Unclassified Metazoa</td>
<td>4.8</td>
<td>6  Fungi*</td>
<td>4.3</td>
</tr>
<tr>
<td>7  Fungi*</td>
<td>4.6</td>
<td>7  Unclassified Metazoan</td>
<td>4.1</td>
</tr>
<tr>
<td>8  Dinoflagellates*</td>
<td>4.4</td>
<td>8  Protist – Algae</td>
<td>2.8</td>
</tr>
<tr>
<td>9  Protist - Algae</td>
<td>2.2</td>
<td>9  Protist - Unclassified</td>
<td>2.4</td>
</tr>
<tr>
<td>10 Protist – Unclassified</td>
<td>2.2</td>
<td>10 Green Algae*</td>
<td>2.1</td>
</tr>
<tr>
<td>11 Protist – Fungi Like</td>
<td>1.2</td>
<td>11 Protist – Fungi Like</td>
<td>2.0</td>
</tr>
<tr>
<td>12 Viridiplantae</td>
<td>0.86</td>
<td>12 Viridiplantae</td>
<td>1.4</td>
</tr>
<tr>
<td>13 Macroalgae</td>
<td>0.24</td>
<td>13 Animalia</td>
<td>1.0</td>
</tr>
<tr>
<td>14 Animalia</td>
<td>0.05</td>
<td>14 Macroalgae</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Table S3: Taxonomic definitions of the six most taxa groups used in the risk assessment.

<table>
<thead>
<tr>
<th>eDNA Benthic Taxa Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diatom</strong></td>
</tr>
<tr>
<td>Kingdom - Protist, Phylum - Bacillariophyta.</td>
</tr>
<tr>
<td><strong>Dinoflagellates</strong></td>
</tr>
<tr>
<td>Kingdom - Protist, Phylum - Dinoflagellata.</td>
</tr>
<tr>
<td><strong>Green Algae</strong></td>
</tr>
<tr>
<td>Photosynthesizing algae-like organisms from the Viridiplantae Kingdom. Phylums - Chlorophyta and Phragmoplastophyta.</td>
</tr>
<tr>
<td><strong>Fungi</strong></td>
</tr>
<tr>
<td>Fungi Kingdom. Most abundant phylums are Ascomycota and Basidiomycota.</td>
</tr>
<tr>
<td><strong>Meiofauna</strong></td>
</tr>
<tr>
<td>Sediment dwelling invertebrates from the Kingdom - Metazoa. Most abundant Phylums - Nematoda, Gastrotricha, Arthropoda, Annelida, and Platyhelminthes.</td>
</tr>
<tr>
<td><strong>Protist - Protozoans</strong></td>
</tr>
<tr>
<td>Protist Kingdom -animal-like protists, predominantly non-filamentous and heterotrophic. Most abundant Phylums - Cercozoa, Ciliophoro, Foraminifera and Apusozoa.</td>
</tr>
</tbody>
</table>

S2.2 Determining Water Quality Predictors for the eDNA Endpoints

A combination of analyses and information was used to determine the predictors of benthic groups including vector fitting, the BEST test, simple correlations, and via trial and error of comparing predicted richness to observed richness.

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 variables 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 that 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. I 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 I used it as a predictor in the BN model, it did not accurately predict the benthic endpoint richness. So in this case, I 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 eDNA 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.
S3. RESULTS

S3.1 Big Picture eDNA Analysis – Ordination with NMDS

Ordination of the MOTU data using NMDS conducted using the Jaccard similarly metric for the Summer 2010 and Fall 2012 data (Figure S2). 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 S2a). 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 S2b). 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.
Figure S2. 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 S3). In 2010, the water quality variables TN, TP, Turbidity, Salinity, and DO were the best predictors (p< 0.001) (Figure S3b). 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 S3a).

In 2012, the water quality variables TN, TP, Chl-a, and Salinity were the best predictors (p<0.05) (Figure S3b). Again, the TN and TP gradient was strongest in the direction of the Logan. The Chl-a endpoint also appears to be predictive of the Logan. The Salinity gradient, again, does not points in the direction of a specific estuary (Figure S3b).
Figure S3. 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.
S3.3 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 predictor 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 default eutrophication 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 S4: Summary of analysis used to determined predictors for the BN-RRM model and the final predictors selected.

<table>
<thead>
<tr>
<th>Taxa Group</th>
<th>Vector Fitting (for MOTUs)</th>
<th>Best Test Results (for taxa group richness)</th>
<th>Correlation Results (for taxa group richness)</th>
<th>Priors from Literature (for taxa group richness)</th>
<th>Predictors used in the BN-RRM model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diatom</td>
<td>2010 MOTUs: TN, TP, Turbidity, Salinity</td>
<td>Salinity, Turbidity (cor = 0.35)</td>
<td>TN ($R^2 = 0.25$)*</td>
<td>Salinity</td>
<td>TN, TP, and Salinity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TP ($R^2 = 0.27$)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turbidity ($R^2 = 0.16$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salinity ($R^2 = 0.35$)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dinoflagellate</td>
<td>2012 MOTUs: TN, TP, Salinity</td>
<td>Salinity (cor = 0.07)</td>
<td>TN ($R^2 = 0.006$)</td>
<td>TN</td>
<td>TN, TP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TP ($R^2 = 0.0004$)</td>
<td>TP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turbidity ($R^2 = 0.001$)</td>
<td>Turbidity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salinity ($R^2 = 0.002$)</td>
<td>Salinity</td>
<td></td>
</tr>
<tr>
<td>Green Algae</td>
<td>TN, Salinity (cor = 0.17)</td>
<td>TN ($R^2 = 0.19$)*</td>
<td>TN ($R^2 = 0.032$)</td>
<td>TN</td>
<td>TN, TP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turbidity ($R^2 = 0.022$)</td>
<td>Turbidity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salinity ($R^2 = 0.1$)</td>
<td>Salinity</td>
<td></td>
</tr>
<tr>
<td>Fungi</td>
<td>Salinity (cor = 0.3)</td>
<td>TN ($R^2 = 0.15$)</td>
<td>TN ($R^2 = 0.23$)*</td>
<td>TN</td>
<td>TN, TP, and Salinity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turbidity ($R^2 = 0.14$)</td>
<td>Turbidity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salinity ($R^2 = 0.29$)*</td>
<td>Salinity</td>
<td></td>
</tr>
<tr>
<td>Meiofauna</td>
<td>TN (cor = 0.11)</td>
<td>TN ($R^2 = 0.04$)</td>
<td>TN ($R^2 = 0.07$)</td>
<td>Salinity</td>
<td>TN, TP, and Salinity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turbidity ($R^2 = 0.05$)</td>
<td>Turbidity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salinity ($R^2 = 0.06$)</td>
<td>Salinity</td>
<td></td>
</tr>
<tr>
<td>Protozoan</td>
<td>Salinity (cor = 0.08)</td>
<td>TN ($R^2 = 0$)</td>
<td>TN ($R^2 = 0$)</td>
<td>Salinity</td>
<td>TN, TP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TP ($R^2 = 0.0008$)</td>
<td>Turbidity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turbidity ($R^2 = 0.01$)</td>
<td>Salinity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salinity ($R^2 = 0.002$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure S4: 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.
Figure S5: Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Noosa Estuary with the Lower Noosa risk region stressors selected.
**Middle Noosa**

![Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Noosa Estuary with the Middle Noosa risk region stressors selected.]

**Figure S6:** Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Noosa Estuary with the **Middle Noosa** risk region stressors selected.
Figure S7: Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Noosa Estuary with the Upper Noosa risk region stressors selected.
Figure S8: Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Pine Estuary with the Lower Pine risk region stressors selected.
Figure S9: Bayesian network relative risk method model (BN-RRM) as shown in Netica (Norsys 2014) for the Pine Estuary with the Middle Pine risk region stressors selected.