

**A Quantitative Basis for the Use of Fish as Indicators of River Health  
in Eastern Australia**

By

**Mark J. Kennard**

Bachelor of Science

Master of Philosophy

Centre for Riverine Landscapes, Faculty of Environmental Sciences,  
Griffith University, Brisbane, Australia.

A thesis submitted in fulfilment of the requirements of the degree of  
Doctor of Philosophy

2005

## Synopsis

In response to increasing community concern with declines in the ability of aquatic ecosystems to deliver critical goods, services and long-term benefits to human society, there is greater recognition of the need for quantitative procedures to assess aquatic ecosystem health and monitor biotic responses to remedial management. This thesis aims to evaluate the potential to incorporate attributes of freshwater fish assemblages into an ecosystem health monitoring program for wadeable rivers and streams in coastal catchments of south-eastern Queensland, Australia.

I identify five key requirements of a quantitative and defensible river health assessment program that need to be evaluated before indicators based on fish can be validly applied for river health assessment in the region. The five requirements are: 1) quantification of error associated with sampling fish; 2) assessment of natural ranges of spatial and temporal variation in fish assemblage attributes; 3) accurate definition of the reference condition expected for these attributes in the absence of human disturbance; 4) demonstrated relationships of the indicators with disturbance; and 5) evaluation of potentially important confounding environmental and biological factors. These are critical considerations for minimising the frequency of Type I errors (incorrectly classifying a site as impaired) and Type II errors (incorrectly classifying a site as unimpaired) and accurately assessing river health.

The ability to develop an efficient data sampling program without sacrificing accuracy and precision, and hence ability to detect changes through space and time, is a critical requirement of every river health assessment program. I demonstrated that accurate and precise reach-scale estimates of fish assemblage attributes, such as species richness, species composition and species relative abundances, could be obtained from multiple-pass electrofishing plus seine netting of three mesohabitat units and concluded that this was generally a more efficient sampling protocol than less intensive sampling over larger spatial scales. I demonstrated that relatively small differences (e.g. < 20%) in these assemblage attributes could be detected with a high statistical power ( $1-\beta > 0.95$ ) and that relatively few stream reaches (e.g. < 4) needed to be sampled to accurately estimate assemblage attributes within close proximity (e.g. 20% of the half width of the confidence interval) to the true population means.

Fish assemblage attributes that respond to anthropogenic disturbances but exhibit low natural temporal variability are potentially the most sensitive yet robust indicators of human impacts for use in bioassessment programs. I showed that some characteristics of fish assemblages (e.g. species richness, composition and to a lesser extent, species relative abundances) were generally highly stable through time, both seasonally and inter-annually. In contrast, fish assemblages occurring in streams with highly variable flow regimes were more variable. This variability appeared due to natural impacts associated with low flow disturbance. However, fish assemblages at these sites appeared resilient to these natural disturbances, provided that flow and habitat conditions resembled the pre-disturbance state.

A critical underpinning of bioassessment programs is the ability to accurately define attributes of the assemblage at a location that are expected in the absence of anthropogenic disturbance (i.e. defining the reference condition). I developed and validated a multivariate predictive model to define the reference condition for native fish species composition based on a large set of least-disturbed reference sites. I also compared four separate approaches to defining the reference condition for native species richness, a univariate biotic attribute underpinning many river health assessment programs.

I demonstrated in that spatial variation in these attributes of fish assemblages could be predicted using a small number of simple environmental variables, but that the accuracy of predictions varied with the method used. The multivariate model developed for sites sampled on one occasion and in one season only (winter), was able to accurately predict fish assemblage composition at sites sampled during other seasons and years, provided that they were not subject to unusually extreme environmental conditions. I showed that the predictive ability of the 'maximum species richness line' (MSRL), an approach commonly used for defining the reference condition for univariate attributes such as native species richness, is inherently compromised by the scoring procedure used and because it is limited to predicting variation in the dependent variable along a single environmental gradient (e.g. stream size). This resulted in substantially higher prediction error in comparison to three alternative regression methods, which incorporate single or multiple environmental variables as predictors. The MSRL approach led to an increased frequency of Type I errors for least-disturbed reference and

validation sites (i.e. incorrectly classing these sites as disturbed) and compromised the ability of the metric to accurately detect a disturbance ‘signal’ at the set of notionally disturbed test sites.

Assessment of the relationships of the fish indicators with human disturbance gradients indicated that streams affected by human activity due to catchment land use and associated local riparian, in-stream habitat and water quality degradation are more likely to be susceptible to invasion by alien fish species and display major differences in native fish assemblage composition and native species richness from that expected by comparison with similar reference areas minimally affected by human disturbances.

Provided that the key requirements of river health assessment identified in this thesis are satisfied, I conclude that indicators based on native fish assemblage composition, native species richness and alien fish species are potentially powerful indicators of human disturbance and can form the basis for a river health monitoring program in south-eastern Queensland and other similarly variable environments in Australia and elsewhere.

## **Declaration**

I hereby declare that this work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where due reference is made in the thesis itself.

Mark J. Kennard

## Acknowledgments

Brad Pusey provided me with the inspiration to attempt this study. I thank him for his supervision throughout this project and for sharing with me over many years his friendship, extensive knowledge, critical mind and enthusiasm for fish ecology. My principal supervisor, Angela Arthington, was instrumental in making this study possible. I thank her for her unwavering support, advice, critical feedback and constant encouragement. Thanks also to my supervisor, Bronwyn Harch, for sharing her breadth of knowledge on statistics and sampling design, and for making it seem interesting and fun (I never thought I'd say that).

Many people helped with field work and made the countless long hours in the field a enjoyable experience. Thanks to Steve Mackay in particular, and also to Peter Benson, Nick Cilento, Elli Dore', Ashley Druve, John Esdaile, Nick Marsh, Darren Renouf, Mick Smith, Wendy Neilan, Celia Thompson, Chris Thompson and Michelle Wood. I thank David Allan, Michael Arthur, Stuart Bunn, Phil Cassey, Rick Cunjak, Lorann Downer, Kurt Fausch, Wade Hadwen, Simon Linke, Nick Marsh, Chris Marshall, Jon Marshall, Steve Mackay, Richard Norris, Julian Olden, LeRoy Poff, Tarmo Raadik, Wayne Robinson and Mick Smith for various helpful discussions on aspects of this thesis. I thank Mick Smith and Christy Fellows for providing me with water quality data. Thanks to Stuart Bunn and all my colleagues in the Centre for Riverine Landscapes for providing an intellectually stimulating workplace. Thanks also to Deslie Smith, Lacey Shaw and Maria Barrett for their cheerful and expert administrative support.

Funding for sampling of reference sites was funded largely by the former Land and Water Resources Research and Development Corporation (LWRRDC) and the former Queensland Departments of Natural Resources and Primary Industries (Fisheries). Funding for sampling of test sites was provided by the Moreton Bay Waterways and Catchments Partnership. Funding support was also provided by the Cooperative Research Centre for Freshwater Ecology, the Centre for Riverine Landscapes and the School of Australian Environmental Studies, Faculty of Environmental Sciences, Griffith University.

I would like to acknowledge several people for their contributions to the publications arising from this thesis. Brad Pusey and Angela Arthington contributed to study design and interpretation of results. Bronwyn Harch gave expert advice on data analysis and interpretation of results. Stephen Mackay contributed to field sampling, data entry and discussed many aspects of stream ecology during our long days in the field together. Elli Dore' contributed to field sampling, data entry and discussion of electrofishing protocols. All made editorial improvements to the manuscripts.

This PhD research project was approved by the Ethics Committee for Experimentation on Animals, Griffith University, and the applied research protocols used were conducted in accordance with the requirements of this Committee.

I would like to thank my family, particularly Jill and Colin for encouraging me always, and for supporting me materially, intellectually and emotionally.

Special thanks to my lovely Lorann for the wonderful life we share together. Thankyou also for introducing me to the warm fuzzy love of two dachshunds, Maxim and Sasá. I dedicate this thesis to the three of you.

<b>Table of Contents .....</b>	<b>Page</b>
<b>Synopsis.....</b>	<b>i</b>
<b>Declaration.....</b>	<b>iv</b>
<b>Acknowledgments.....</b>	<b>v</b>
<b>Table of Contents.....</b>	<b>vi</b>
<b>List of Tables.....</b>	<b>xi</b>
<b>List of Figures.....</b>	<b>xv</b>
<b>Chapter 1: General introduction.....</b>	<b>1</b>
<b>1.1. Background.....</b>	<b>1</b>
<i>1.1.1. Models of ecosystem structure and function.....</i>	<i>1</i>
<i>1.1.2. Human impacts on aquatic ecosystems.....</i>	<i>3</i>
<i>1.1.3. Assessment of aquatic ecosystem ‘health’.....</i>	<i>4</i>
<i>1.1.4. Indicators of ecosystem health.....</i>	<i>5</i>
<i>1.1.5. Approaches to ecosystem health assessment using biota.....</i>	<i>7</i>
<b>1.2. Aims and structure of thesis.....</b>	<b>10</b>
<b>Chapter 2: Study area.....</b>	<b>14</b>
<b>2.1. Location.....</b>	<b>14</b>
<b>2.2. Climate and hydrology.....</b>	<b>14</b>
<b>2.3. Land use and human impacts.....</b>	<b>17</b>
<b>2.4. Data sets.....</b>	<b>18</b>
<b>Chapter 3: Estimating local stream fish assemblage attributes: sampling effort and efficiency at two spatial scales.....</b>	<b>20</b>
<b>3.1. Synopsis.....</b>	<b>20</b>
<b>3.2. Introduction.....</b>	<b>21</b>
<b>3.3. Methods.....</b>	<b>23</b>
<i>3.3.1. Study area and sampling sites.....</i>	<i>23</i>
<i>3.3.2. Fish sampling procedures.....</i>	<i>25</i>
<i>3.3.3. Habitat sampling procedures.....</i>	<i>27</i>
<i>3.3.4. Data analysis.....</i>	<i>28</i>
<i>3.3.4.1. Sampling effort and efficiency within mesohabitat units.....</i>	<i>29</i>
<i>3.3.4.2. Sampling effort and efficiency at the stream reach scale.....</i>	<i>31</i>

<b>3.4. Results.....</b>	<b>33</b>
<b>3.4.1. <i>Sampling effort and efficiency within mesohabitat units.....</i></b>	<b>33</b>
<b>3.4.2. <i>Interspecific variation in sampling efficiency and mesohabitat use.....</i></b>	<b>37</b>
<b>3.4.3. <i>Sampling effort and efficiency at the stream reach scale.....</i></b>	<b>38</b>
<b>3.4.4. <i>Accumulation of habitat.....</i></b>	<b>45</b>
<b>3.5. Discussion.....</b>	<b>47</b>

<b>Chapter 4: Stability, persistence and resilience of freshwater fish assemblages     across gradients of flow and habitat variability.....</b>	<b>53</b>
<b>4.1. Synopsis.....</b>	<b>53</b>
<b>4.2. Introduction.....</b>	<b>55</b>
<b>4.3. Methods.....</b>	<b>59</b>
<b>4.3.1. <i>Hydrology and sampling sites.....</i></b>	<b>59</b>
<b>4.3.2. <i>Fish sampling procedures.....</i></b>	<b>62</b>
<b>4.3.3. <i>Choice of environmental variables.....</i></b>	<b>63</b>
<b>4.3.4. <i>Data analysis.....</i></b>	<b>65</b>
4.3.4.1. <i>Characterisation of environmental variability.....</i>	65
4.3.4.2. <i>Characterisation of fish assemblage stability and persistence.....</i>	66
4.3.4.3. <i>Relationships between environmental variability, fish assemblage                 variability and potential confounding factors.....</i>	69
4.3.4.4. <i>Characterisation of fish assemblage resilience.....</i>	69
<b>4.4 Results.....</b>	<b>71</b>
<b>4.4.1. <i>Environmental variability template.....</i></b>	<b>71</b>
<b>4.4.2. <i>Fish assemblage variability and relationships with environmental variability                 and potential confounding factors.....</i></b>	<b>73</b>
<b>4.4.3. <i>Resilience.....</i></b>	<b>76</b>
<b>4.5. Discussion.....</b>	<b>78</b>
<b>4.5.1. <i>Stability and persistence.....</i></b>	<b>78</b>
<b>4.5.2. <i>Covariates and confounding factors.....</i></b>	<b>79</b>
<b>4.5.3. <i>Relationships of fish assemblage variability with environmental                 variability.....</i></b>	<b>80</b>
<b>4.5.4. <i>Mechanisms enabling resilience to disturbance.....</i></b>	<b>81</b>
<b>4.5.5. <i>Implications of temporal variability for river health assessment.....</i></b>	<b>83</b>

<b>Chapter 5: Development and application of a predictive model of freshwater fish assemblage composition to evaluate river health.....</b>	<b>86</b>
5.1. Synopsis.....	86
5.2. Introduction.....	87
5.3. Methods.....	89
5.3.1. <i>Reference, validation and test site data sets</i> .....	89
5.3.2. <i>Fish sampling procedures</i> .....	93
5.3.3. <i>Estimation of environmental variables and characterisation of human disturbance gradient</i> .....	94
5.3.4. <i>Statistical methods</i> .....	96
5.3.4.1. <i>Development of predictive model</i> .....	96
5.3.4.2. <i>Model validation and the effect of temporal variation on model performance</i> .....	98
5.3.4.3. <i>Model testing and sensitivity to disturbance</i> .....	98
5.4. Results.....	99
5.4.1. <i>Freshwater fish fauna and biological characteristics of the reference data set</i> .....	99
5.4.2. <i>Model development: classification and discriminant analysis</i> .....	99
5.4.3. <i>Internal consistency of predictive model</i> .....	100
5.4.4. <i>Validation of predictive model and effect of temporal variation on model accuracy</i> .....	104
5.4.5. <i>Testing the ability of the predictive model to detect human disturbance</i> ...	108
5.5. Discussion.....	111
<b>Chapter 6: Accurately defining the reference condition for summary biotic metrics: a comparison of four approaches.....</b>	<b>118</b>
6.1. Synopsis.....	118
6.2. Introduction.....	120
6.3. Methods.....	122
6.3.1. <i>Reference, validation and test site data sets</i> .....	122
6.3.2. <i>Fish sampling procedures</i> .....	123
6.3.3. <i>Estimation of environmental variables and characterisation of disturbance</i> .....	123
6.3.4. <i>Statistical methodology</i> .....	124
6.3.4.1. <i>Modelling approaches used to define the reference condition</i> .....	124

6.3.4.2. <i>Maximum Species Richness Line (MSRL)</i> .....	125
6.3.4.3. <i>Simple Linear Regression (SLR)</i> .....	125
6.3.4.4. <i>Generalised Linear Modelling (GLM)</i> .....	126
6.3.4.5. <i>Regression Tree Modelling (TREE)</i> .....	126
6.3.4.6. <i>Validation and testing of models: direct comparison of model accuracy, error rates and ability to characterise the disturbance gradient</i> .....	127
<b>6.4. Results</b> .....	<b>129</b>
<b>6.4.1. Defining the reference condition</b> .....	<b>129</b>
6.4.1.1. <i>Maximum species richness line (MSRL)</i> .....	129
6.4.1.2. <i>Simple Linear Regression (SLR)</i> .....	129
6.4.1.3. <i>Generalised Linear Modelling (GLM)</i> .....	131
6.4.1.4. <i>Regression tree modelling (TREE)</i> .....	132
<b>6.4.2. Direct comparison of model error rates and ability to characterise the disturbance gradient at test sites</b> .....	<b>134</b>
<b>6.5. Discussion</b> .....	<b>137</b>
<b>6.5.1. Metric scoring and Type I errors - Comparisons between MSRL and SLR</b> .....	<b>137</b>
<b>6.5.2. Single versus multiple environmental gradients – comparisons between SLR, GLM and TREE</b> .....	<b>139</b>
<b>6.5.3. Implications for assessment of river health</b> .....	<b>140</b>
 <b>Chapter 7: Are alien fish a reliable indicator of river health?</b> .....	<b>142</b>
<b>7.1. Summary</b> .....	<b>142</b>
<b>7.2. Introduction</b> .....	<b>144</b>
<b>7.3. Methods</b> .....	<b>147</b>
7.3.1. <i>Data sets</i> .....	147
7.3.2. <i>Fish sampling procedures</i> .....	148
7.3.3. <i>Estimation of environmental variables and characterisation of disturbance</i> .....	148
7.3.4. <i>Statistical methods</i> .....	148
<b>7.4. Results</b> .....	<b>151</b>
7.4.1. <i>Characteristics of the fish fauna</i> .....	151
7.4.2. <i>Relationships of alien fish with sample effort, catchment variables and site physical characteristics</i> .....	154
7.4.3. <i>Relationships of alien fish with the native fish fauna</i> .....	154

7.4.4. <i>Relationships of alien fish with indicators of human disturbance</i> .....	158
7.5. Discussion.....	160
<b>Chapter 8: General discussion</b> .....	<b>165</b>
8.1. The need for quantitative methods to assess river health.....	165
8.2. River health assessments are only as good as the data that underpins them.....	166
8.3. Biotic assemblages vary through space and time, but is variation natural or human induced? .....	167
8.4. Procedures for defining the reference condition: should accuracy be sacrificed for simplicity? .....	170
8.5. Detecting disturbance and diagnosing causes? .....	173
8.6. Assessing river health using fish in south-eastern Queensland.....	175
<b>Chapter 9: References</b> .....	<b>177</b>
<b>Appendix 1: Published paper arising from this thesis</b> .....	<b>206</b>

<b>List of Tables.....</b>	<b>Page</b>
<b>Table 3.1.</b> Summary (means, standard deviations and ranges) of physical characteristics and water quality parameters in mesohabitats sampled in the Mary and Albert Rivers (n = 72 samples). .....	25
<b>Table 3.2.</b> Characteristics of fish species collected from mesohabitats in the Mary River and Albert River. Maximum sizes (cm standard length) are those commonly attained by each species in south-eastern Queensland and behavioural designations refer to those species that are benthic (B), pelagic (P) or frequently inhabit submerged vegetation (SV) (Pusey <i>et al.</i> 2004, M. J. Kennard, unpublished data). The relative abundance and frequency of occurrence of each species are expressed as percentages of the total number of fishes collected and mesohabitats sampled (n = 18,341 individual and 72 samples, respectively). Native fish families are arranged in approximate phylogenetic order (after Pusey <i>et al.</i> 2004) and alien species are denoted by *. All fish illustrations by B.J. Pusey except <i>C. carpio</i> (Pedro T. Escudero, Mindanao State University, Philippines) and <i>G. holbrooki</i> and <i>X. helleri</i> (sourced from <a href="http://www.dpi.qld.gov.au/extra/aquaticinvaders">www.dpi.qld.gov.au/extra/aquaticinvaders</a> ). .....	30
<b>Table 3.3.</b> Spearman's rank correlation coefficients between measures of sampling effort (total time spent sampling and percentage of total time spent sampling on electrofishing pass 1) and physical, chemical and biological attributes of the study sites. All mesohabitat samples were included in analyses (n = 72 samples). Correlation coefficients underlined are significant at $P < 0.05$ after correction for multiple comparisons using the Dunn-Sidak procedure (Quinn and Keough, 2002). .....	35
<b>Table 3.4.</b> <i>F</i> values (Wilks Lambda) and their associated significance levels for repeated measures Analysis of Variance of within-mesohabitat type (cumulative passes) and between-mesohabitat type variation in log(X+1) transformed estimates of within-mesohabitat total species richness, total abundance, species presence/absence and species relative abundances. See text for explanation of contrasts. ....	37
<b>Table 3.5.</b> Relative sampling efficiency (efficiency ratio) of single pass electrofishing versus multiple pass electrofishing plus seine netting to achieve a desired accuracy (80% and 90%) and precision (0.2 and 0.1) in estimating total species richness, species composition and species relative abundance in a river reach. The number of mesohabitat units required to achieve the prescribed accuracy and precision levels were obtained from Figure 3.6. Total mean sampling duration was estimated by multiplying the number of mesohabitat units by the sampling time required by a two or three person crew to collect, identify and enumerate fish using single pass electrofishing (mean 23 minutes per mesohabitat) or multiple (three) pass electrofishing (mean 45 minutes) plus seine netting (estimated as 10 minutes per site). Efficiency ratios were calculated by dividing the total mean sampling duration of single pass electrofishing by the duration of multiple pass electrofishing plus seine netting. Note that single pass electrofishing did not estimate species richness or relative abundance to an accuracy of 90% after the maximum of 6 sampled mesohabitats (Fig. 3.5) so efficiency ratios could not be calculated. ....	46

**Table 4.1.** Median and range of values for landscape, hydrologic and hydraulic variables used to define the environmental variability template for sites in the Mary River (n = 16 locations) and Albert River (n = 11 locations). Hydrologic variability characteristics were calculated for the four-year period (June 1993 to July 1997) leading up to and including the sampling period (June 1994 – July 1997). Hydraulic variability characteristics were calculated from instantaneous measures collected on each seasonal sampling occasion (maximum of 10 samples). CV refers to coefficient of variation and ARI refers to the annual return interval flood. The variable loadings on each of the first four components of a principal components analysis are also shown (see also Fig. 4.4). The percentage variation explained by each principal component is given in parentheses and the highest variable loadings on each principal component are shown in bold type. ....64

**Table 4.2.** Pearson's correlation coefficients for relationships between indices of biological variability (based on all species and individuals), environmental variability and biological co-variates at 27 sites in the Mary and Albert Rivers, south-eastern Queensland sampled on 7 – 10 occasions between 1994 and 1997. Indices of environmental variability are the principal component scores from the analysis presented in Table 4.1 and Figure 4.4. Significant correlations are shown in bold type ( $p < 0.01$  after correction for multiple comparisons using the Dunn-Sidak procedure (Quinn and Keough, 2002)). Correlations between fish assemblage calculated before and after excluding rare and uncommon species are also shown (underlined), all except mean CV of individual species biomass were significant at  $p < 0.01$ . ....74

**Table 4.3.** Mean spearman's rank correlation coefficients ( $\pm$  SD) and number of significant correlations for relationships between years in the rank order of magnitude of hydraulic characteristics and fish assemblage attributes (species richness, total abundance and total biomass) among sites. Mean correlation coefficients for between-year relationships in the ranked abundance and biomass of each species pooled over all study sites each for annual sample are also shown. Analyses were conducted for every possible between-year comparison for each season (winter, spring and summer). Correlations were corrected for multiple comparisons using the Dunn-Sidak procedure (Quinn and Keough, 2002) with a significance level of  $p < 0.01$ . ....77

**Table 5.1.** Range and median values of environmental predictor variables at reference, validation and test sites. N indicates the number of sites for catchment-related variables that were static in time. For environmental variables that varied through time (at validation sites sampled on multiple occasions), N refers to the number of samples (indicated by \*). ....93

**Table 5.2.** Variables used to quantify the source and intensity of anthropogenic disturbance at each site. Variables describe human land use, water chemistry, riparian vegetation and in-stream habitat characteristics. A description of each variable, the expected generalised response to disturbance (in parentheses) and the range of values observed at the test sites are given. Numbers in superscript refer to the methods used to measure each variable. ....95

**Table 5.3.** Mean density (number of fish.100m<sup>-2</sup>) ( $\pm$  SE) and frequency of occurrence (percentage of sites) of the 24 fish species within each reference site group defined by UPGMA classification. Also shown are mean values ( $\pm$  SE) and F values for environmental variables identified by stepwise multiple discriminant analysis as being significant predictors of UPGMA classification group membership ( $p < 0.001$ ). .....101

**Table 5.4.** Means and standard deviations of the expected number of species (sum of the predicted species probabilities of occurrence), the observed number of species and Observed/Expected ratios for reference sites used to construct the model and calculated using probability of occurrence thresholds of  $PO > 0\%$  and  $PO > 50\%$ . The width of the reference band (90<sup>th</sup> – 10<sup>th</sup> percentile of all O/E scores) for each PO threshold is given in parentheses. Means and standard deviations for species metrics at reference sites within each classification group are also shown. ....102

**Table 5.5.** Results of Friedman's test of concordance (*FR*) and Kendall's coefficient of concordance (*KCC*) in  $OE_{50}$  scores at internal validation sites and external validation sites in the Mary and Albert Rivers. Numbers of sites in each data set and river are given in parentheses. Also shown are F statistics and p values for one-way repeated measures analysis of variance values to test for trends in  $OE_{50}$  scores through time. ..107

**Table 5.6.** Principal components analysis of 16 disturbance variables from the 47 test sites in south-eastern Queensland. The percentage variation explained by each component is given in parentheses and the highest variable loadings on each principal component are shown in bold type. ....109

**Table 5.7.** Summary of multiple regression model to predict variation in  $O/E_{50}$  scores at the 48 test sites according to variation in the disturbance gradient variables (principal components). The approximate model  $R^2$  and the relative importance of each predictor variable fitted in the model (indicated by the percent of total variance explained) is given. ....110

**Table 6.1.** Pearson's *r* correlation coefficients for relationships between environmental variables at reference sites. Variables transformed ( $\log_{10}(x+1)$ ) prior to analysis are denoted by <sup>L</sup>. Correlations significant at  $p < 0.05$  after correction for multiple comparisons using the Dunn-Sidak procedure (Quinn & Keough, 2002) are denoted by \*. Also shown are  $R^2$  values for Simple Linear Regression relationships between species richness and individual environmental variables (significance levels for each relationship are given in parentheses). ....130

**Table 6.2.** Summary of Generalised Linear Model for predicting species richness values at the 62 reference sites according to catchment and local site descriptors. The relative importance of each predictor variable fitted in the model is indicated by the sums of squares and the percent of total variance explained (approximate model  $R^2 = 69\%$ ,  $p < 0.0001$ ). The  $C_p$  statistic indicates the terms included in the model (those greater than the Null  $C_p$  value). ....132

**Table 6.3.** Comparison of the ability of the four prediction methods to accurately characterise the disturbance gradient using multiple regression models to predict variation in species richness O/P scores at the 48 test sites according to variation in the disturbance gradient variables (principal components). For each GLM model the approximate model  $R^2$  and the relative importance of each predictor variable fitted in the model (indicated by the percent of total variance explained) is given. All models were significant at  $p < 0.001$ . ....136

**Table 7.1.** Native and introduced freshwater fish species present in south-eastern Queensland (including all coastal catchments from the Burnett River south to the border with New South Wales). Species native to the region are denoted by N, translocated native species by T (? denotes species of uncertain origin) and alien species introduced from other countries and thought to have established by A (source: Pusey *et al.* 2004). Species collected during the present study are denoted by \*. .....152

**Table 7.2.** Spearman's rank correlation coefficients for relationships between alien fish species indices (abundance and biomass) and variables describing sampling effort and landscape and local scale environmental data at the 47 sites sampled. Also shown are significant rank correlations between alien fish indices and disturbance gradient variables. Correlations significant at  $p < 0.05$  after correction for multiple comparisons using the Dunn-Sidak procedure (Quinn and Keough, 2002) are denoted by \*. .....155

**Table 7.3.** Summary of multiple regression models to predict variation in alien species relative abundance and biomass at the 47 test sites according to variation in the disturbance gradient principal components. For each GLM model the approximate model  $R^2$  and the relative importance of each predictor variable fitted in the model (indicated by the percent of total variance explained) is given. ....159

## List of Figures.....Page

**Figure 1.1.** Simplified conceptual model of linkages between abiotic drivers and biotic responses in functional aquatic ecosystems that deliver goods, services and long-term benefits for human society (linkages shown with dark solid arrows). Human disturbances impact natural aquatic ecosystems and hence affect the societal benefits they provide (linkages shown with dashed open arrows). The nature and strength of these relationships is contingent on the spatial and temporal scale of examination. Figure modified from Lorenz *et al.* (1997) and Baron *et al.* (2002), using information sourced from Allan & Flecker (1993). .....2

**Figure 2.1.** Map of study area showing location of reference sites and test sites in river basins of south-eastern Queensland (Note: some sites located close together overlay each other and hence may not be visible on map). Major impoundments and place names mentioned in the text are also depicted. The insets show the location of the study area in Queensland, Australia. ....15

**Figure 2.2.** Plots of mean daily minimum temperature (closed squares), mean daily maximum temperature (open squares) and mean monthly rainfall (open circles) for each month at Gympie (central Mary River basin) and Beaudesert (central Logan-Albert River basin), south-eastern Queensland. Data record length was >80 years for both locations. Data source: Queensland Bureau of Meteorology, figure modified from Pusey *et al.* (2004). ....16

**Figure 2.3.** Variation in mean total monthly discharge for gauging stations in the lower Mary River and lower Albert River, south-eastern Queensland. The catchment area ( $\text{km}^2$ ) upstream of each gauging station is given in parentheses. Data record length was >20 years for each gauge. Data source: Queensland Department of Natural Resources and Mines, figure modified from Pusey *et al.* (2004). ....17

**Figure 3.1.** Spatial scale of sampling in the Mary River and Albert River in south-eastern Queensland, Australia. (a) Location of stream reaches where sampling of mesohabitat units was undertaken. (b) The subset of stream reaches where extended sampling of multiple (six) mesohabitat units was undertaken in the Mary River is depicted in (a) by numbered squares. (c) Sampling points within each mesohabitat unit where measurements of in-stream and bank habitat were taken. ....24

**Figure 3.2.** Mean ( $\pm$  SE) cumulative percent of total sampling duration (elapsed time and electrofisher power on time) spent on each consecutive electrofishing pass at individual mesohabitat units in the Mary and Albert Rivers. Inset box plots show the variation in total elapsed time and power on time spent sampling individual mesohabitat types. The lines at the top, middle and bottom of each box represent the 75<sup>th</sup> percentile, median and 25<sup>th</sup> percentile, respectively. Upper and lower bars represent 90<sup>th</sup> and 10<sup>th</sup> percentiles and mean values are represented by squares. ....34

**Figure 3.3.** Changes in estimates of fish species richness, total abundance, species presence/absence and species relative abundance with increasing sampling effort within riffles (circles), runs (triangles) and pools (squares). For total species richness and total abundance, the data represents the sequential increase in mean ( $\pm$  SE) cumulative percentage of the total catch collected by each electrofishing pass and supplementary seine-netting. Inset box plots show the variation in total species richness and total abundance collected within each mesohabitat type. For species presence/absence and species relative abundance, the data represents the mean ( $\pm$  SE) Bray-Curtis similarity between each cumulative pass and the total assemblage (derived from all electrofishing passes plus seine-netting). .....36

**Figure 3.4.** Interspecific variation in (a) sampling efficiency and (b) mesohabitat use. Sampling efficiency is defined as the probability of capture (as assessed by mean frequency of occurrence in mesohabitat samples) and percentage of total catch (mean relative abundance) for each species (grouped by behavioural type) sampled by single pass electrofishing, multiple pass electrofishing and supplementary seine netting. Mesohabitat use is indicated by the mean frequency of occurrence and mean relative abundance of each species occurring in riffles, runs or pools. Sample sizes for each species (number of individuals or number of mesohabitats) can be calculated using the data presented in Table 3.2). .....39

**Figure 3.5.** Fish species relative abundance (percentage of total abundance) versus frequency of occurrence (percentage of mesohabitats occupied) averaged across the six Mary River stream reaches sampled. Each point represents one species sampled using single pass electrofishing or multiple pass electrofishing plus seine netting (species discussed in the text are denoted by the first letter of the genus and species, respectively). Data are plotted on log scale. Rare species are defined as forming less than 10% of total abundance and discontinuous species are defined as occurring in less than 10% of mesohabitats (averaged across stream reaches). .....40

**Figure 3.6.** Mean cumulative species richness versus length of stream sampled using (a) single-pass electrofishing and (b) multiple-pass electrofishing plus seine netting in each of the six stream reaches. Reach numbers match those given in Figure 3.1. Estimates of mean species richness for each sampling interval (cumulative number of mesohabitats) were based on 1000 bootstrapped randomisations that incrementally constructed hypothetical series of mesohabitat units. For clarity, standard deviations are not shown. .....41

**Figure 3.7.** Change in average sampling accuracy ( $\pm$  SD) and average precision versus length of stream sampled using (a) single-pass electrofishing and (b) multiple-pass electrofishing plus seine netting to estimate total fish species richness (circles), species composition (triangles) and species relative abundance (squares). Accuracy for each cumulative number of mesohabitats sampled is represented by the mean ( $\pm$  SD) percentage of total fish species richness (closed circles) or mean ( $\pm$  SD) Bray Curtis similarity with total species composition (closed squares) and species relative abundance (closed triangles), respectively. Precision for each cumulative number of mesohabitats sampled is represented by the coefficient of variation (mean/SD). Estimates of accuracy and precision were based on 1000 bootstrapped randomisations that incrementally constructed hypothetical series of mesohabitat units. Results are averaged across six sampled stream reaches in the Mary River. .....43

**Figure 3.8.** (a) The power ( $1-\beta$ ) of different fish sampling methods to detect a 20% decrease ( $\delta$ ) in estimates of total species richness (circles), species composition (triangles) and species relative abundance (squares), with increasing number of mesohabitats sampled ( $\alpha = 0.05$ ). (b) The number of stream reaches required (based on sampling of three mesohabitats per reach) to achieve a half width of the confidence interval within given percentages of the true population mean for total species richness, species composition and species relative abundance ( $\alpha = 0.05$ ). Data from single pass electrofishing is shown with open symbols and dashed lines, data from multiple-pass electrofishing plus seine netting is shown with closed symbols and solid lines. ....44

**Figure 3.9.** Cumulative increases in the mean percentage ( $\pm$  SD) of the total number of habitat types encountered in stream reaches versus length of stream sampled (expressed as number of mean stream widths). Separate curves are shown for depth, velocity and substrate categories, and for bank microhabitat and in-stream microhabitat categories. Means were generated using 1000 bootstrapped randomisations that incrementally constructed hypothetical combinations of habitat categories from observed samples and averaged over the six stream reaches. Frequency histograms of the number of habitat sampling points within each habitat category used in the analyses are also shown. ....47

**Figure 4.1.** Variation in daily river discharge ( $\text{ML}\cdot\text{day}^{-1}$ ) for Glastonbury Creek in the Mary River (see Figure 4.2 for location) during the period leading up to and including the study period. Arrows highlight key aspects of the flow regime of potential ecological importance for fish, and hypothesised to cause temporal variation in species composition, abundance and biomass. Also shown are the timing of fish and habitat surveys undertaken in winter, spring and summer between 1994 and 1997. ....57

**Figure 4.2.** Location of the study sites (depicted by circles) in the Mary and Logan-Albert River basins in south-eastern Queensland, Australia. Site symbols are colour-coded and sized according to a gradient of relative hydrologic and hydraulic variability determined by a principal components analysis (see Table 4.1 and Figure 4.4). Negative loadings on the first principal component (depicted by larger open circles) indicate increasingly stable environmental conditions, positive loadings (larger closed circles) indicate increasingly variable environmental conditions. Glastonbury Creek, the site used to depict flow regime variation in Figure 4.1, is marked with \*. ....60

**Figure 4.3.** Box plots showing spatial variation in mean monthly discharge characteristics at 16 sites in the Mary River and 11 sites in the Albert River over a 25 year period (1<sup>st</sup> January 1974 – 31<sup>st</sup> December 1998). The plots show (a) mean daily runoff ( $\text{ML}\cdot\text{Km}^2$ ), (b) mean coefficient of variation of daily discharge (%), (c) mean number of flood events per year (flood events  $> 1.67$  year annual return interval) and (d) mean number of zero flow events per year. The lines at the top, middle and bottom of each box represent the 75<sup>th</sup> percentile, median and 25<sup>th</sup> percentile, respectively. Upper and lower bars represent 90<sup>th</sup> and 10<sup>th</sup> percentiles and means are represented by symbols. ....61

**Figure 4.4.** Spatial variation in Mary River and Albert River sites across an environmental variability template described by a principal components analysis of landscape, hydrologic and hydraulic variables (PC1 vs PC3). Principal component 1 (50% of the variance) describes a gradient of low flow and hydraulic variability and principal component 3 (11% of the variance) describes a gradient of daily flow variability and baseflow contribution. Variable loadings on each component are given in Table 1. ....72

**Figure 4.5.** Cumulative frequency distributions of total variability in fish assemblage attributes during the study period for sites sampled in the Mary River (closed squares) and Albert River (closed circles). Also shown are cumulative frequency distributions of annual variability in fish assemblage attributes for comparisons of winter samples (open squares), spring samples (open diamonds) and summer samples (open triangles) in the Mary River. See text for explanation of each fish assemblage variability index. ....75

**Figure 5.1.** Location of reference sites (numbered by classification group membership), validation sites and test sites in south-eastern Queensland (Note: some sites located close together overlay each other and hence may not be visible on map). Major impoundments are also depicted. The inset shows the location of the study area in Queensland, Australia. ....90

**Figure 5.2.** Frequency distributions of observed over expected ratios for the 72 reference sites at probability of occurrence thresholds of  $PO > 0\%$  (open bars) and  $PO > 50\%$  (closed bars). ....102

**Figure 5.3.** Relationship between expected and observed values for (a)  $PO_0$  and (b)  $PO_{50}$  probability of occurrence thresholds at reference sites. The diagonal dashed lines represent the line of perfect agreement between the two measures. Regression lines and equations for each plot are also shown. For each  $PO$  threshold,  $p > 0.05$  for  $H_0$ : slope = 1 and  $H_0$ : Intercept = 0. Also shown are the relationships between the total number of species observed at reference sites and (c)  $O/E_0$  and (d)  $O/E_{50}$  scores. For each  $PO$  threshold,  $p < 0.001$  for  $H_0$ : slope = 0. ....104

**Figure 5.4.** Relationship between expected and observed values for (a) internal validation samples and (b) external validation samples at the  $PO_{50}$  probability of occurrence threshold. The five internal validation sites sampled during spring 2000 are shown as black diamonds (note that two sites directly overlie each other on the plot). The diagonal dashed lines represent the line of perfect agreement between the two measures. Regression lines and equations for each plot are also shown. For each dataset,  $p > 0.05$  for  $H_0$ : slope = 1 and  $H_0$ : Intercept = 0. ....105

**Figure 5.5.** Temporal variation in  $O/E_{50}$  scores for (a) internal validation samples and (b) external validation samples in the Mary River (open circles) and Albert River (closed circles). The 90<sup>th</sup> – 10<sup>th</sup> percentile reference band is depicted with horizontal dashed lines. Internal validation sites sampled in Winter 1994 (Figure 5.5a) were included in the reference data set used to construct the model and were used to derive  $O/E_{50}$  scores for these samples, all other samples were foreign to the model. ....106

**Figure 5.6.** Frequency distributions of  $O/E_{50}$  ratios for (a) reference, internal validation and external validation sites and (b) reference and test sites). Vertical dashed lines indicated the width of the reference band. ....108

**Figure 5.7.** Difference in mean values ( $\pm$  SE) of disturbance variables significantly different at sites where each species was predicted (at the  $PO_{50}$  level) and observed to occur (open bars), sites where each species was predicted to occur but was not observed (closed bars), and sites where each species was not predicted to occur but was observed (hatched bars). Chi-square values for Kruskal Wallance tests are given for each comparison; all were significant at  $p < 0.05$  after correction for multiple comparisons using the Dunn-Sidak procedure (Quinn and Keough, 2002). Sample sizes for each site category are shown below the x-axis of each plot. ....111

**Figure 6.1.** Species richness versus  $\log_{10}$  catchment area for reference sites. The maximum species richness line (MSRL), trisections used for metric scoring and the simple linear regression line (SLR) are also shown. Validation and test sites are also shown on plot. Note that sites overlaying each other may be obscured on the plot. ...130

**Figure 6.2.** Box plots of variation in the magnitude of the residuals (difference between values predicted and those observed) from each prediction method using each environmental predictor variable for reference and validation sites. The lines at the top, middle and bottom of each box represent the 75<sup>th</sup> percentile, median and 25<sup>th</sup> percentile, respectively. Upper and lower bars represent 90<sup>th</sup> and 10<sup>th</sup> percentiles and means are represented by symbols. ....131

**Figure 6.3.** Species richness predicted by the GLM model versus observed values for (a) the reference sites and validation sites and (b) the reference sites and test sites. Standard errors around predicted values for validation and test sites are also shown. ....132

**Figure 6.4.** Regression tree for predicting species richness at reference sites, environmental variables used in forming the tree and their critical values for determining the splits. The distribution of values for species richness in each reference site group is given in the box plot. The number of sites within each reference site group is given in parentheses. The predicted group membership (using the regression tree) and observed values for species richness at validation (open squares) and test sites (closed triangles) are also shown. ....133

**Figure 6.5.** Comparison of residual mean squared error (RMSE) scores for each prediction method and each environmental predictor variable for reference and validation sites. ....134

**Figure 6.6.** Frequency (%) of reference, validation and test sites failing pre-defined biocriteria for each prediction method. Biocriteria were defined on the basis of the MSRL trisection method (those sites scoring less than 5 fail) or species richness Observed/Predicted scores (O/P scores <75% fail) (note that elevation was used as the environmental predictor variable for the MSRL and SLR methods). ....135

**Figure 7.1.** Biological characteristics of the fish fauna at the study sites showing (a) relative abundance (closed bars, scale on left axis) (n=4943 individuals) and frequency of occurrence (open bars, scale on right axis) (n=47 sites), and (b) relative biomass (total of 69.1 kg). Alien species are denoted by A and translocated native species by T, in parentheses. Inset histograms show the relative abundance (n=1041 individuals) and relative biomass (total of 1.8 kg) of alien species only (species names are abbreviated to the first letter of the genus and species, respectively). ....153

**Figure 7.2.** Relationships of (a) alien species relative abundance and (b) alien species relative biomass with native species richness observed at the study sites. Also shown are relationships of alien species relative abundance (c) and relative biomass (d) with total densities and biomass densities of native species. Spearman's rank correlation coefficients and their significance levels for each plot are also shown. Note that some axes are plotted on a  $\log_{10}(x+1)$  scale for clarity. ....156

**Figure 7.3.** Relationships of (a) alien species relative abundance and (b) alien species relative biomass with the ratio of observed to predicted native species richness at the study sites. Sites considered to be of high biological integrity (O/P scores >0.75, n = 27 sites) are denoted by open circles, and those of low biological integrity (O/P scores <0.75, n = 20 sites) by closed triangles. Also shown are relationships of alien species relative abundance (c) and relative biomass (d) with predicted native species richness. The same symbols are used for each site type. Spearman's rank correlation coefficients and their significance levels for each plot are also shown. Note that the relative biomass data are plotted on a  $\log_{10}(x+1)$  scale for clarity. ....157

**Figure 7.4.** Mean ( $\pm$  SE) values of disturbance variables significantly different at sites where alien species were present (closed bars, n=28 sites) and absent (open bars, n=19 sites). Mann-Whitney test statistics ( $T$ ) are given for each comparison; all were significant at  $p < 0.05$  after correction for multiple comparisons using the Dunn-Sidak procedure (Quinn and Keough, 2002). ....158

**Figure 7.5.** Relationships between alien species relative abundance (a) and relative biomass (b) predicted by multiple regression models and the values observed at each site. The disturbance variables used as predictors for each model are given in Table 3. Dashed lines indicate hypothetical 1:1 relationships between predicted and observed values. The approximate model  $R^2$  is given in parentheses for each plot. ....160