A fish-based Estuarine Health Index for the Swan Estuary, WA

Chris Hallett

Centre for Fish and Fisheries Research, School of Biological Sciences and Biotechnology, Murdoch University, South Street, Murdoch, Western Australia 6150, Australia.
Tel: +61 8 92398802; fax: +61 8 92398899; e-mail: c.hallett@murdoch.edu.au

Abstract

I describe the development of a fish-based, multimetric estuarine health index (EHI) for the Swan Estuary, Western Australia. A suite of fish community characteristics (metrics), including measures of species composition, diversity and abundance, trophic structure and life history function, were selected via a novel weight of evidence approach, on the basis of their sensitivity to detect inter-annual change in estuarine condition. Seasonally-adjusted reference conditions for each selected metric were established for each region of the Swan Estuary using 30 years' of historical fish assemblage data, and thus represent a best available standard of biotic integrity against which the current and future health of the estuary may be assessed and compared. Scores for each metric were assigned according to the extent of the metric's deviation from its reference condition. Values for the EHI were calculated from summed metric scores for each main region of the Swan Estuary for each season and year, to identify trends in the recent health of the estuary, and to validate the sensitivity and reliability of the index. The index, which is the first such tool to be developed for Western Australia, will provide managers with a reliable and cost-effective, quantitative method for assessing and communicating the health of the Swan Estuary.

Keywords: Biotic integrity, ecological indicator, multimetric index,

Introduction

Multimetric indices integrate information from a broad suite of characteristics of the biological communities upon which they are based, to provide an assessment of the ecological integrity ('health') of ecosystems. Such indices have become a key component of estuarine monitoring and management programs in the United States, South Africa and Europe (Deegan et al. 1997; Bilkovic et al. 2005; Harrison and Whitfield 2006; Roset et al. 2007; Borja et al. 2008) yet, to date, their development and application to Australian estuaries has been limited (Deeley and Paling 1998; Scheltinga and Moss 2007).

Estuaries in south-western Australia are increasingly subjected to numerous stressors, with several of these systems being extensively modified by human activities and only one, Broke Inlet, having been assessed as near-pristine during a recent broad-scale national assessment of estuarine status (NLWRA 2002). Many of the stressors affecting these estuaries are exemplified in arguably the most intensively impacted and best-studied estuary of south-western Australia, the Swan-Canning Estuary (hereafter referred to as the Swan Estuary; Fig. 1).

Extensive land clearance for urban and agricultural development within its catchment has greatly increased the magnitude of stressors acting upon the Swan Estuary since European settlement during the early to mid 1800s. These include increased delivery of sediments and nutrients to estuarine waters, leading to persistent eutrophication (Hamilton and Turner 2001) and mounting salinisation, extending the spatial and temporal persistence of vertical stratification and hypoxic conditions within the estuary (Hamilton et al. 2001). In response to these stressors, the Swan Estuary is displaying signs of a general decline in ecosystem health, particularly in its upper reaches (Swan River Trust 1999; 2003, 2009b) The system now regularly suffers from periods of severe hypoxia (Douglas et al. 1997) and phytoplankton blooms, including those of toxic species (Hosja and Deeley 1994), which have resulted in several large fish mortalities.

Despite these problems, resource managers of estuaries in Western Australia currently lack a reliable, simple and affordable method for (i) quantifying the environmental health of estuaries relative to appropriate reference conditions, (ii) monitoring temporal changes in estuarine health to detect deterioration beyond critical thresholds and (iii) detecting those regions of individual estuaries at greatest risk of environmental decline. This project thus aimed to develop and validate a fish-based, multimetric estuarine health index (EHI) for the Swan Estuary, Western Australia, to address these needs.

Methods

Metric selection

An extensive range of candidate fish community characteristics (metrics) were tested for their suitability for incorporation into the EHI, including measures of species composition, diversity and abundance, trophic structure (*i.e.* feeding relationships) and functional aspects of the assemblage, *e.g.* life history functions such as the degree to which various taxa use estuaries as spawning or nursery areas (see Table 1 for a refined list of these metrics).

A novel, objective, approach was employed for selecting that subset of candidate metrics which most consistently reflected temporal changes at the ecosystem level, and thus which are likely to be most sensitive to changes in ecosystem condition. A combination of multivariate statistical analyses and information-theoretic techniques were employed to assess both metric redundancy and sensitivity, enabling responsive and parsimonious subsets of candidate metrics to be selected according to the weight of evidence from multiple analyses of data sets collected using divergent sampling techniques over various historical sampling periods. Selected metrics were subsequently incorporated into indices for assessing the health of the nearshore waters (< 2 m depth) and offshore waters (> 2 m depth) of this system. A full and detailed account of the selection of metrics for the EHI is provided in Hallett *et al.* (in prep.).

Establishing reference conditions and scoring metrics

Numerous sets of historical fish species abundance data collected from the nearshore waters of the Swan Estuary since 1976, including those from the current study, were employed in determining reference conditions for each of the selected nearshore fish metrics. However, given the divergent nearshore sampling methods employed in the Swan

Estuary between 1976 and 2009, the resulting data sets are each affected by differing biases, preventing them from being directly comparable. In order to reliably incorporate the collective nearshore fish assemblage data to establish reference conditions for each of the selected nearshore fish metrics, a net comparison study was carried out (Hallett 2010) and equivalence factors derived for standardising the abundances of fish species across the three main nearshore sampling methods employed historically (namely 21.5, 41.5 and 133 m seine nets).

Count data for each of the fish species encountered during the net comparison study were summed by habitat guild and subjected to negative binomial regression analysis to assess the influence of net type on fish counts (Hallett 2010). For each of the five habitat guilds, equivalence factors (β) were derived from the best model by exponentiation of the statistically significant ($P < 0.05$) estimates of the parameter coefficients for the 41.5 and 133 m nets, and 95% confidence intervals on those equivalence factors were determined as $\exp(\beta \pm 2 \times \text{SE})$ (Maki *et al.* 2006). The appropriate equivalence factors were then applied to all counts of fish species in samples collected using the 41.5 and 102-133 m nets, to obtain an aggregated data set in which all samples were standardised to counts per 21.5 m net. Where no significant effect of a given net on a specific guild was identified, the original count data were left unadjusted.

In contrast to the historical fish assemblage studies carried out in the nearshore waters of the Swan Estuary, those undertaken in the offshore waters have employed relatively consistent methods and effort, such that they are largely free from sampling bias. The historical and current fish abundance data obtained from the offshore waters throughout the estuary using gill nets were thus collated for use in determining reference conditions for each of the selected offshore fish metrics, following their conversion to catch hr^{-1}.

Reference conditions for each selected nearshore and offshore metric were determined by identifying the best available value recorded during any of the fish faunal studies carried out between 1976 and 2009 in each of those waters. Identification of these 'best' values for each metric (*i.e.* whether they were among the lowest or highest of all values ever recorded) depended on whether the predicted response of the metric was positive or negative. In addition to the methodological biases noted above, the historical and current nearshore data sets were also biased by the influence of seasonal and regional differences in fish community structure. Therefore, to eliminate the potential for these biases to impact the reliability of reference conditions, best available reference values for each selected nearshore or offshore metric were established for each region*season combination using all available current and historical data, which had been standardised for net type where necessary. Given that the current study aimed to develop an index of estuarine health to aid in the future management of the Swan Estuary, the regions for which specific reference conditions were established were redefined according to the Ecological Management Zones recently established for the system by the Swan River Trust (2009a; Fig. 2).

Values for each of the selected fish metrics were calculated from the standardised data for each historical and current fish sample. The appropriate region*season-specific reference conditions for each metric were then defined statistically from these metric data and used to establish metric scores for each sample via continuous scaling, as outlined by Minns *et al.* (1994), Hughes *et al.* (1998) and Hering *et al.* (2006). Upper and lower thresholds were set using percentiles, rather than minima and maxima, to avoid the influence of extreme outliers (Gibson *et al.* 2000), and scores between these upper and lower thresholds were calculated by linear interpolation. In cases where metric values exceeded the best threshold, a metric score of 10 was allocated. Moreover, when no fish were caught in a sample, all metrics received a score of zero.

Index calculation and validation

Index scores for both the nearshore and offshore health indices were calculated by summing the scores for their component metrics, then adjusting the resultant value by the number of metrics in the index to produce a final index score that ranged from 0-100 (Ganasan and Hughes 1998). Index scores were calculated for each historical and current sample and were then averaged to provide a measure of the health of the Swan Estuary in each of the years in which fish were sampled between 1976 and 2009.

Index scores were then used to determine thresholds for establishing qualitative estuarine health status by subdividing their possible range into four equal classes (Table 2). It was considered that more classes than this would make decisions regarding management actions more problematic (Ganasan and Hughes 1998; Qadir and Malik 2009), whilst fewer classes might allow the health of the estuary to decline markedly before a health status threshold is crossed and management actions are indicated.

Whilst index scores enable preliminary interpretation of spatial and/or temporal trends in estuarine health, a thorough evaluation of index reliability is required before the index may be applied as a management tool. Cross-validation approaches were thus employed to quantify the sources of variability in the nearshore and offshore index scores and thus assess their reliability. Data collected only during the current study were used for these analyses as, compared to the various historical studies, the data sets from this period were collected across all regions of the estuary in eight consecutive seasons, and thus were the most comprehensive and consistently recorded. To address the question of whether the variability of index scores among groups of replicate sites differed between regions and/or seasons, the standard deviations of the scores for each of those groups were calculated and compared. To determine whether inter-seasonal variation in index scores at a site was related to the ecological quality of that site, the standard deviation of the index scores among seasons in each year was plotted against the corresponding mean index score for each site. Spearman's correlation test was used to determine if ρ, calculated between the standard deviations and the means of the scores, differed significantly from zero at $P = 0.05$. The extent of the variability in index scores between consecutive years, and thus the effects of this variability on the consistency of health status classifications, was also determined by plotting index scores from sites assessed in each season in 2007/08 against those from the same sites assessed in the same seasons of 2008/09. Spearman's correlation test was used to determine if ρ, calculated between the scores from the first *vs* the second of the above years, differed significantly from zero at $P = 0.05$.

Finally, bootstrap cross-validation was employed to quantify the effects of random sampling variability on index scores, as described for the IBI by Fore *et al.* (1994) and

Dolph *et al.* (2010). One thousand bootstrap samples were created for each fish sample collected during the current study by randomly resampling from the original sample data with replacement (Efron and Tibshirani 1993). An index score was calculated for each bootstrap sample and the mean of these scores was calculated for each site on each sampling occasion. The percentile method (Efron and Tibshirani 1993) was used to estimate a 95% confidence interval for the index score, and the length of the confidence interval was determined by calculating the difference between the upper and lower confidence limits (Dolph *et al.* 2010). The bias of the index was quantified for each site visit by subtracting the original index score from the mean bootstrapped score (Fore *et al.* 1994). The effects of sampling variability on the consistency of health status classifications were also investigated by determining the proportion of samples for which the health status indicated by the mean bootstrapped score differed from that indicated by the original index score.

Results

Metric selection

Sets of 11 and seven metrics were selected for inclusion in nearshore and offshore indices of estuarine health, respectively (Table 3).

Establishing reference conditions and scoring metrics

The reference conditions for each nearshore metric, as determined from the best available metric values derived from the standardised seine net data collected between 1976 and 2009, are presented for each region*season combination in Table 4. For several of these metrics, there were clear differences in reference condition values both between different regions in a given season, and between seasons within a region. For example, the reference condition for the metric *No species* varied from as few as five species in the Upper Swan Estuary in winter, to as many as 14 species in the Canning Estuary/Lower Canning River in summer or the Middle Swan Estuary in summer or autumn. The reference condition values for each of the offshore metrics are presented in Table 5 and, like those for the nearshore metrics, clear differences occurred both between regions in each season and vice versa.

Index calculation and validation

Examination of the changes in mean nearshore index scores between 1976 and 2009 suggests that the health of the nearshore waters of the Swan Estuary has remained relatively constant over the last three decades, with the health status being classified as fair throughout this time (Fig. 3). However, it is important to note that reliable interpretation of the longer-term trends in these mean scores is impeded by differences among studies in the location, timing and intensity of sampling, as well as by the inability to standardise values of species richness among samples collected using different net types. Changes in nearshore index scores from 2005 to 2009 (and, to a lesser extent, those from 1995 to 2001) may, however, be interpreted reliably, due to greater standardisation of the sampling methodology across this period. Although the lack of sampling in consecutive years between 1993/94 and 2003/04 reduces the ability to discern index trends over that time, there is evidence to suggest that the health of the nearshore waters of the estuary has increased in more recent years, from a mean health index score of *ca* 58 in 2005/06 to 64 in 2008/09 (Fig. 3).

In contrast, the longer-term trends in the health of the offshore waters of the Swan Estuary may be interpreted reliably, due to the greater consistency of sampling methodologies among all historical and current fish community studies of those waters. The mean offshore index score has decreased consistently from 56.5 in the late 1970s to 47 in 2008/09, resulting in the health status of these waters being classified as poor during the most recent study period, for the first time in three decades (Fig. 4).

In both years of the current study, between-site variability of the nearshore index within any given season was, on average, lower in the more upstream regions of the Swan Estuary (Middle-Downstream to Upper Swan River) than in those regions nearer the mouth of the system (Channel, Basin and Canning River: Fig. 5). A similar pattern was also observed in the degree to which the standard deviations of index scores varied among seasons, with that in the upstream regions often being considerably less pronounced than in regions further downstream, most notably in 2008/09 (Fig. 5b). Across all regions, the seasons with the lowest variability of index scores (*i.e.* those with the most points below the average standard deviation) were summer and autumn in 2007/08 and summer and winter in the following year (Fig. 5).

Variability of index scores among replicate sites within a region was generally greater for the offshore index than for its nearshore equivalent (*cf.* Figs 5 and 6). Also, unlike the nearshore index, between-site variability of offshore index scores generally decreased in a downstream direction during autumn and particularly winter in 2007/08 (Fig. 6a), while the same was often true during winter in 2008/09 (Fig. 6b). The variability of offshore index scores was lowest, on average, in spring and summer in 2007/08 and in autumn during 2008/09.

The standard deviations of index scores among seasons at each nearshore site sampled between 2007 and 2009 exhibited a weak, negative correlation with the means of those scores, which was close to being statistically significant (ρ = -0.246, P = 0.056; Fig. 7). Thus, inter-seasonal variation in nearshore index scores at the various sites in any given year was largely unrelated to the ecological quality of those sites. Moreover, when the samples in Fig. 7 were coded for region, there was no evidence to suggest that seasonal variability in index scores at a site in any given year was related to the region of the estuary in which the site was located.

In contrast, significant and moderate negative correlations were observed between the inter-seasonal variation in index scores and the averages of those scores at each offshore site in 2007-09 (ρ = -0.553, P < 0.001; Fig. 8). These results thus demonstrated that inter-seasonal variation in offshore index scores was inversely related to site quality. Most notably, sites in the Upper Swan Estuary (USE) generally had lower ecological quality and, in accordance with the above significant relationship, greater seasonal variability of index scores, than did sites from other regions (Fig. 8).

Nearshore index scores recorded at each site in each season of the first year of the current study were significantly, yet weakly, positively correlated with those for the corresponding samples in the second year (ρ = 0.211, P = 0.027). The small extent of this correlation suggests that there were often considerable differences in index scores between the two years. Nonetheless, this inter-annual variability had a relatively minor impact on the consistency of health status classifications, as most nearshore sites were assessed as good/fair in both years. Index scores from offshore sites also exhibited a weak, positive correlation between years, although this was not significant (ρ = 0.224, P = 0.059). Inter-annual variability in offshore index scores had a greater impact on the consistency of health status classifications than in the case of the nearshore index.

The length of the 95% confidence intervals (CI) around bootstrapped mean nearshore index scores ranged from zero to approximately 27 points, with a mean of seven points. The bias of the original nearshore index scores ranged from one point (underestimation) to approximately -7 points (overestimation), with a mean negative bias of one to two points. Original index scores thus consistently overestimated estuarine health, most notably among higher quality nearshore sites (Fig. 9). However, for only 16 out of 233 site visits (approximately 7%) did the difference between the mean bootstrap score and the original index score represent a change in health status classification.

In the case of the offshore index, the length of the 95% CIs around mean bootstrapped scores ranged from zero to 40 points, with a mean of approximately 14 points. The bias of original offshore index scores ranged from a 12 point underestimation to an overestimation of approximately -30 points, with a mean bias of *ca* -4 points. Original index scores of < 45 thus represented probable underestimates of estuarine health, but those at the higher end of the index scale tended to overestimate health (Fig. 10). The difference between the mean bootstrap score and the original index score represented a change in health status classification for 31 out of 119 site visits (*i.e.* 26%), of which two-thirds were overestimates, *i.e.* the original index score indicated a higher health status than did the bootstrap score.

Discussion

Metric selection

Multimetric biotic indices derived using an objective, statistical approach to metric selection are widely regarded as being more robust than those in which metric selection is based on expert judgement alone (Hering *et al.* 2006; Roset *et al.* 2007). The multifaceted statistical approach employed in the current study succeeded in objectively selecting that combination of fish metrics which is potentially best able to reflect inter-annual changes in the environmental condition of the Swan Estuary. A potential weakness of the current approach is that further work is required to quantify the sensitivity of metrics to specific anthropogenic stressors. Nonetheless, the consistent decrease observed in offshore health index scores over the last three decades suggests that this index is capable of detecting the widely-perceived, long-term decline in the condition of the offshore waters of the Swan Estuary (Swan River Trust 1999; 2003; Valesini *et al.* 2005).

The respective sets of 11 and seven metrics selected for the nearshore and offshore waters of this system represented a range of fish community characteristics including species composition and diversity, trophic structure, life history and habitat functions and, in the case of the nearshore index, a potential sentinel species. Biotic indices constructed from a broad range of metrics such as this are more likely to reflect the integrated ecological effects of multiple and diverse stressors, and thus reveal their impacts on the condition of the estuary as a whole (Barbour *et al.* 1995). Although six of the seven metrics selected for the offshore index were common to the nearshore index, the requirement for separate indices applicable to nearshore and offshore waters stems mainly from difficulties in standardising sampling effort across the very different fish sampling techniques employed in these two water depths (seine and gill nets, respectively).

Establishing reference conditions and scoring metrics

Several authors have highlighted problems commonly associated with the use of historical data for establishing reference conditions, including a lack of quantity or quality of data and a lack of standardised methods for data collection (Hughes 1995; Harrison and Whitfield 2004). In the case of the nearshore index presented here, the combined historical data set used to establish reference conditions comprised almost 2,000 samples collected from throughout the Swan Estuary over three decades. Although the quantity of data available was thus acceptable, the methodological differences in its collection, which necessitated the application of complex data standardisation procedures (Hallett 2010), compromised the quality of that data to some extent. In contrast, standardisation of sampling methodology and intensity across the historical and current studies of the offshore fish fauna in the Swan Estuary enabled the reliable integration of those data for setting reference conditions.

For each of the selected nearshore and offshore fish metrics, appropriate reference conditions were defined statistically for each region*season combination to eliminate the potential for spatial and temporal biases to impact the reliability of reference conditions (Karr 1999; Kennard *et al.* 2006a; Coates *et al.* 2007). It is widely recommended that appropriate reference conditions for ecological integrity metrics be established from a population of carefully-selected, minimally-impaired reference sites that are located across multiple systems subject to differing levels of human stress and identified using independent measures of environmental quality (Hughes 1995; Gibson *et al.* 2000; USEPA 2006). However, the present study attempted to develop a multimetric index for a single estuarine system, without access to an established, independent means of identifying minimally-impacted sites or gradients of anthropogenic disturbance. Consequently, biological reference conditions were instead defined from the 'best' fraction of the observed metric values across a large number of sites throughout the system (Gibson *et al.* 2000; Blocksom 2003), thus enabling the future health of the system and the success of its management to be measured in terms of deviation from this best available reference state.

Index calculation and validation

Average health index scores for the nearshore waters of the Swan Estuary appear to have undergone a moderate degree of variation from 1976 to 2009, although the resulting health status has remained fair throughout this time. This suggests that the above classification is likely to be robust to natural variability over longer time scales. More detailed examination of the trends in nearshore index values should, however, be undertaken with caution at this stage, as the lack of methodological consistency between the various nearshore fish community studies presents problems for index interpretation. Thus, whilst the equivalence factors derived during the current study appear to provide a satisfactory means of adjusting fish densities to account for differences in methodological bias between the various seine nets employed from the late 1970s to 2009, they do not enable the adjustment of species richness (Hallett 2010). Consequently, those fish metrics based on numbers of species in a sample will remain subject to the bias associated with the net used to obtain the sample. As the 133 m seine nets were employed exclusively between 1976 and 1982, whereas sampling in subsequent years has employed the more comparable 41.5 and 21.5 m seines, it is important to note that inter-annual changes in nearshore index scores can be interpreted with greater confidence among the more contemporary studies (*i.e.* those from the mid-1990s onwards). These latter studies also benefit from greater consistency in the location, timing and intensity of sampling, compared to that of the earliest survey period.

In contrast, inter-annual changes in the health of the offshore waters of the Swan Estuary may be interpreted more reliably, due to the largely consistent sampling methodology and intensity employed among all studies of these waters. The fact that mean offshore index scores have decreased over the last three decades, and that their health status is now classed as poor, as opposed to fair during all previous studies, indicates that the ecological health of the deeper waters of the estuary has declined over this time. Given that trends in the nearshore index since the mid-2000s indicate the opposite, it is suggested that these findings reflect a movement of the fish community inhabiting deeper waters toward nearshore habitats.

Differences in the variability of index scores among replicate sites were identified between both regions and seasons. Within any given season, nearshore index scores were less spatially variable in the more upstream regions of the Swan Estuary in both years of the current study, which is possibly explained by the reduced habitat heterogeneity of these regions compared to that of regions nearer the mouth of the system. Variability of index scores among replicate sites was generally greater in the offshore than nearshore waters, which largely reflected a greater prevalence of zero fish catches in offshore samples, and most notably in the upstream regions in winter. It should be noted that the measure of index variability employed was strongly affected by zero catches, as the standard deviations of index scores in each of the regions and seasons were calculated from only three replicate site visits. If more sites were sampled within each region, and/or if sites were sampled more regularly, it would be possible to determine whether such zero catches were more likely to be anomalous (false zeros; Cunningham and Lindenmayer 2005), or reflective of a genuine tendency across the region or season towards low index scores in a given period (true zeros). Further work is thus needed to quantify the effects of sampling intensity within a region and season on the precision of the health indices, and thus to determine the optimum spatio-temporal level of sampling required for a robust future monitoring regime in the Swan Estuary. Given that the between-site variability of the nearshore and offshore indices was most consistently low during summer and autumn, it is suggested that the optimum sampling period for applying these indices in the Swan Estuary is from December to May. However, the variability of index scores within seasons must also be examined before an optimum sampling period can be definitively identified (Yoder and Rankin 1995).

No evidence was observed of a relationship between the ecological quality of nearshore sites and the inter-seasonal variability of their index scores. Moreover, inter-seasonal index variability did not exhibit obvious differences among regions. Such findings parallel those of Pyron *et al.* (2008). In contrast, and as reflected in the findings of other authors, the variability of index scores among seasons at offshore sites was inversely related to the quality of those sites, with poorer quality sites exhibiting greater inter-seasonal differences in their index scores than sites of higher quality (Karr *et al.* 1987; Fore *et al.* 1994; Yoder and Rankin 1995; Deegan *et al.* 1997; Bilkovic *et al.* 2005; Brooks *et al.* 2009). Given the above, variability in index scores has therefore been proposed as a signal of ecological degradation, with impacted sites thought to be less resilient to natural temporal changes in abiotic factors (Fore *et al.* 1994; Simon 1999; Paller 2002). The far greater inter-seasonal variability observed among offshore than nearshore index scores may thus be further evidence that the deeper, offshore waters of the Swan Estuary are in poorer health than the nearshore waters of this system. In particular, the low and highly variable index scores for most of the offshore sites in the Upper Swan Estuary support the contention that this region is the most severely impacted (Swan River Trust 1999; 2003).

Inter-annual changes in index scores between the two consecutive years of the current study were relatively large, and were notably greater than those reported by Harris and Silveira (1999) for an IBI applied to rivers in New South Wales. The weak positive correlations between 2007/08 and 2008/09 for both indices highlight the fact that numerous sites exhibited a considerable change in index score between years. Given that such longer-term variation in index scores may, of itself, provide an indication of ecological disturbance (Fore *et al.* 1994; Deegan *et al.* 1997; Hughes *et al.* 1998; Paller 2002), the greater inter-annual variability of the offshore index in the current study may be further evidence of the lower health status of the deeper waters in the Swan Estuary. Nonetheless, it is important to note that, although the inter-annual variability of index scores was relatively high, health status classifications were fairly robust to these changes, as was also found by Pyron *et al.* (2008) for an IBI applied to the Wabash River in Indiana.

Fewer than 25% of nearshore index scores varied by more than 10 points as a result of random sampling error. Thus, the precision of the current nearshore index was higher than that of a fish-based IBI applied to Minnesotan river basins, for which almost 25% of scores varied by 15 or more points (Dolph *et al.* 2010). The precision of the offshore index scores was lower than that of the nearshore index, with the most variable score having a range of 40 points, although this precision was comparable to that documented by Dolph *et al.* (2010) for the IBI.

The mean bias of the nearshore index across all sites was only one or two points, and the difference between the mean bootstrap score and the original index score represented a change in health status classification on only 7% of occasions. Such findings suggest that the method developed for classifying the health of nearshore sites is robust to the effects of random sampling variability. In contrast, the bias of the offshore index resulted in original scores exceeding mean bootstrap scores by 20 or more points in some cases, and indicated the potential for a change in health status classification for 26% of site visits during 2007-09. Given also that the bias of the offshore index is inconsistent, confidence limits around health status thresholds may be appropriate to account for the observed lack of index precision, as have been established for the IBI in some jurisdictions (Gibson *et al.* 2000; Wan *et al.* 2010).

Conclusions

Validation of the nearshore and offshore health indices developed for the Swan Estuary during this study has demonstrated their capability for tracking long-term changes in the perceived health of this system. Moreover, classification of the health status of this system on the basis of index scores was fairly robust to the effects of both natural spatio-temporal variability and sampling error. Thus, these estuarine health indices, which are the first such tool to be developed for estuaries in Western Australia, provide managers with a reliable, practical and cost-effective method for assessing the ecological health of the Swan Estuary, and potentially other estuaries across WA, and a simple, visual method for communicating the health of estuaries to the public and stakeholders.

References

Barbour, M.T., Stribling, J.B., Karr, J.R. (1995). Multimetric approach for establishing biocriteria and measuring biological condition. In: Davis, W.S., Simon, T.P. (Eds.), Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, Florida, pp. 63-77.

Bilkovic, D.M., Hershner, C.H., Berman, M.R., Havens, K.J., Stanhope, D.M. (2005). Evaluating nearshore communities as indicators of ecosystem health. In: Bortone, S.A. (Ed.), Estuarine Indicators. CRC Press, Boca Raton, Florida, pp. 365-379.

Blocksom, K.A. (2003). A performance comparison of metric scoring methods for a multimetric index for mid-Atlantic highlands streams. *Environmental Management* 31, 670-682.

Borja, A., Bricker, S.B., Dauer, D.M., Demetriades, N.T., Ferreira, J.G., Forbes, A.T., Hutchings, P., Jia, X., Kenchington, R., Marques, J.C., Zhu, C. (2008). Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. *Marine Pollution Bulletin* 56, 1519-1537.

Brooks, R., McKenney-Easterling, M., Brinson, M., Rheinhardt, R., Havens, K., O'Brien, D., Bishop, J., Rubbo, J., Armstrong, B., Hite, J. (2009). A Stream-Wetland-Riparian (SWR) index for assessing the condition of aquatic ecosystems in small watersheds along the Atlantic slope of the eastern U.S. *Environmental Monitoring and Assessment* 150, 101-117.

Coates, S., Waugh, A., Anwar, A., Robson, M. (2007). Efficacy of a multi-metric fish index as an analysis tool for the transitional fish component of the Water Framework Directive. *Marine Pollution Bulletin* 55, 225-240.

Cunningham, R.B., Lindenmayer, D.B. (2005). Modelling count data of rare species: some statistical issues. *Ecology* 86, 1135-1142.

Deegan, L.A., Finn, J.T., Ayvazian, S.G., Ryder-Kieffer, C.A., Buonaccorsi, J. (1997). Development and validation of an estuarine biotic integrity index. *Estuaries* 20, 601-617.

Deeley, D.M., Paling, E.I. (1998). Assessing the ecological health of estuaries in southwest Australia. In: McComb, A.J., Davis, J.A. (Eds.), Wetlands for the Future. Gleneagles, Adelaide, pp. 257-271.

Dolph, C.L., Sheshukov, A.Y., Chizinski, C.J., Vondracek, B., Wilson, B. (2010). The Index of Biological Integrity and the bootstrap: Can random sampling error affect stream impairment decisions? *Ecological Indicators* 10, 527-537.

Douglas, G.B., Hamilton, D.P., Gerritse, R., Adeney, J.A., Coad, D.N. (1997). Sediment geochemistry, nutrient fluxes and water quality in the Swan Estuary, WA. In: Davis, R.J. (Ed.), Managing Algal Blooms: Outcomes from the CSIRO's Multi-Divisional Blue-Green Algal Program. CSIRO Land and Water, Canberra, ACT, pp 15-30.

Efron, B., Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman & Hall, New York, 436 pp.

Fore, L.S., Karr, J.R., Conquest, L.L. (1994). Statistical properties of an index of biological integrity used to evaluate water resources. *Canadian Journal of Fisheries and Aquatic Sciences* 51, 1077-1087.

Ganasan, V., Hughes, R.M. (1998). Application of an index of biotic integrity (IBI) to fish assemblages of the rivers Khan and Kshipra (Madhya Pradesh), India. *Freshwater Biology* 40, 367-383.

Gibson, G.R., Bowman, M.L., Gerritsen, J., Snyder, B.D. (2000). Estuarine and coastal marine waters: bioassessment and biocriteria technical guidance. USEPA report 822-B-00-024. Office of Water, Washington DC, 300 pp.

Hallett, C.S. (2010). The development and validation of an estuarine health index using fish community characteristics. PhD thesis, Murdoch University, 207 pp.

Hallett, C.S., Valesini, F.J., Clarke, K.R. (In prep.). A multimetric health index for the Swan Estuary, Western Australia: Selecting metrics in the absence of independent measures of ecological condition.

Hamilton, D.P., Chan, T., Robb, M.S., Pattiaratchi, C.B., Herzfeld, M. (2001). The hydrology of the upper Swan River Estuary with focus on an artificial destratification trial. *Hydrological Processes* 15, 2465-2480.

Hamilton, D.P., Turner, J.V. (2001). Integrating research and management for an urban estuarine system: the Swan-Canning Estuary, Western Australia. *Hydrological Processes* 15, 2383-2385.

Harris, J.H., Silveira, R. (1999). Large-scale assessments of river health using an Index of Biotic Integrity with low-diversity fish communities. *Freshwater Biology* 41, 235-252.

Harrison, T.D., Whitfield, A.K. (2004). A multi-metric fish index to assess the environmental condition of estuaries. *Journal of Fish Biology* 65, 683-710.

Harrison, T.D., Whitfield, A.K. (2006). Application of a multimetric fish index to assess the environmental condition of South African estuaries. *Estuaries and Coasts* 29, 1108-1120.

Hering, D., Feld, C.K., Moog, O., Ofenböck, T. (2006). Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. *Hydrobiologia* 566, 311-324.

Hosja, W., Deeley, D.M. (1994). Harmful phytoplankton surveillance in Western Australia. Waterways Commission Report No. 43, Perth, WA, 99 pp.

Hughes, R.M. (1995). Defining acceptable biological status by comparing with reference conditions. In: Davis, W.S., Simon, T.P. (Eds.), Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, Florida, pp. 31-47.

Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Kincaid, T.M., Reynolds, L., Larsen, D.P. (1998). A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences* 55, 1618-1631.

Karr, J.R. (1999). Defining and measuring river health. *Freshwater Biology* 41, 221-234.

Karr, J.R., Yant, P.R., Fausch, K.D., Schlosser, I.J. (1987). Spatial and temporal variability of the index of biotic integrity in three Midwestern streams. *Transactions of the American Fisheries Society* 116, 1-11.

Kennard, M.J., Harch, B.D., Pusey, B.J., Arthington, A.H. (2006). Accurately defining the reference condition for summary biotic metrics: a comparison of four approaches. *Hydrobiologia* 572, 151-170.

Maki, K.L., Hoenig, J.M., Olney, J.E., Heisey, D.M. (2006). Comparing historical catch rates of American shad in multifilament and monofilament nets: a step toward setting restoration targets for Virginia stocks. *North American Journal of Fisheries Management* 26, 282-288.

Minns, C.K., Cairns, V.W., Randall, R.G., Moore, J.E. (1994). An index of biotic integrity (IBI) for fish assemblages in the littoral zone of Great Lakes areas of concern. *Canadian Journal of Fisheries and Aquatic Sciences* 51, 1804-1822.

NLWRA. (2002). Australia's Natural Resources: 1997-2002 and beyond. NLWRA, Canberra, 163 pp.

Paller, M.H. (2002). Temporal variability in fish assemblages from disturbed and undisturbed streams. *Journal of Aquatic Ecosystem Stress and Recovery* 9, 149-158.

Pyron, M., Lauer, T.E., LeBlanc, D., Weitzel, D., Gammon, J.R. (2008). Temporal and spatial variation in an index of biological integrity for the middle Wabash River, Indiana. *Hydrobiologia* 600, 205-214.

Qadir, A., Malik, R.N. (2009). Assessment of an index of biological integrity (IBI) to quantify the quality of two tributaries of river Chenab, Sialkot, Pakistan. *Hydrobiologia* 621, 127-153.

Roset, N., Grenouillet, G., Goffaux, D., Pont, D., Kestemont, P. (2007). A review of existing fish assemblage indicators and methodologies. *Fisheries Management and Ecology* 14, 393-405.

Scheltinga, D.M., Moss, A. (2007). A framework for assessing the health of coastal waters: a trial of the national set of estuarine, coastal and marine indicators in Queensland. Final report to the NLWRA. Queensland Environmental Protection Agency, Indooroophily, 265 pp.

Simon, T.P. (1999). Introduction: Biological integrity and use of ecological health concepts for application to water resource characterization. In: Simon, T.P. (Ed.), Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities. CRC Press, New York, pp. 3-16.

Swan River Trust. (1999). Swan-Canning Cleanup Program: An action plan to clean up the Swan-Canning Rivers and Estuary. Swan River Trust / Water and Rivers Commission, East Perth, WA, 102 pp.

Swan River Trust. (2003). Swan-Canning Cleaup Program. Action Plan Implementation: 2003. Swan River Trust, East Perth, WA.

Swan River Trust. (2009a). Ecological management zones for the Swan Canning river system. Available at: www.swanrivertrust.wa.gov.au/science/river.Documents/Ecological%20management%20zones%20for%20the%20Swan%20Canning%20river%20system.pdf. [Last accessed April 2010].

Swan River Trust. (2009b). Swan Canning Water Quality Improvement Plan. Swan River Trust, East Perth, WA, 114 pp.

USEPA. (2006). Developing Biological Indicators: Lessons Learned From Mid-Atlantic Streams. USEPA report EPA/903/F-06/001. United States Environmental Protection Agency, Mid-Atlantic Integrated Assessment, Fort Meade, Maryland, 8 pp.

Valesini, F.J., Hoeksema, S.D., Smith, K.A., Hall, N.G., Lenanton, R.C.J., Potter, I.C. (2005). The fish fauna and fishery of the Swan Estuary: A preliminary study of long-term changes and responses to algal blooms. FRDC report. Murdoch University, Perth, WA, 217 pp.

Wan, H., Chizinski, C.J., Dolph, C.L., Vondracek, B., Wilson, B.N. (2010). The impact of rare taxa on a fish index of biotic integrity. *Ecological Indicators* 10, 781-788.

Yoder, C.O., Rankin, E.T. (1995). Biological criteria program development and implementation in Ohio. In: Davis, W.S., Simon, T.P. (Eds.), Biological

Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, Florida, pp. 109-144.

List of Tables

Table 1 Refined list of candidate metrics for possible inclusion in a biotic index of estuarine health for the Swan Estuary.

Metric	**Metric code**	**Metric description**
Species diversity / composition / abundance		
Species richness	*No species*	Total number of species present
Dominance	*Dominance*	No. of species comprising 90% of total individuals
Shannon diversity	*Sh-div*	Shannon's diversity index
Trophic structure		
Proportion of trophic specialists	*Prop trop spec*	Trophic specialists as a proportion of total individuals
Number of trophic specialists	*No trop spec*	Number of trophic specialist species
Proportion of trophic generalists	*Prop trop gen*	Trophic generalists as a proportion of total individuals
Number of trophic generalists	*No trop gen*	Number of trophic generalist species
Proportion of detritivores	*Prop detr*	Detritivores as a proportion of total individuals
Number of detritivores	*No detr*	Number of detritivorous species
Feeding Guild Composition	*Feed guild comp*	Number of different trophic guilds present
Habitat / life history function		
Proportion of benthic species	*Prop benthic*	Benthic associated as a proportion of total individuals
Number of benthic species	*No benthic*	Number of benthic associated species
Proportion of estuarine spawners	*Prop est spawn*	Estuarine spawners as a proportion of total individuals
Number of estuarine spawning species	*No est spawn*	Number of estuarine spawning species
Proportion of estuarine residents	*Prop est res*	Estuarine residents as a proportion of total individuals
Number of estuarine resident species	*No est res*	Number of estuarine resident species
Sentinel species		
Proportion of *P. olorum*	*Prop P. olorum*	*P. olorum* as a proportion of total individuals
Total density of *P. olorum*	*Tot no P. olorum*	Total abundance (density) of *P. olorum*

Table 2 Thresholds for qualitative classification of estuarine health status on the basis of index scores that ranged between 0 and 100.

Index score	**Estuarine health status**
≥ 75	Good
≥ 50 < 75	Fair
≥ 25 < 50	Poor
< 25	Very poor

Table 3 Summary of the fish metrics selected for incorporation into nearshore and offshore indices of estuarine health for the Swan Estuary.

Metric	**Nearshore**	**Offshore**
No species	■	■
Dominance		
Sh-div		■
Prop trop spec	■	
No trop spec	■	■
No trop gen	■	■
Prop detr	■	■
Feed guild comp		
Prop benthic	■	■
No benthic	■	
Prop est spawn	■	■
No est spawn	■	
Prop P. olorum	■	
Tot no P. olorum	■	

Table 4 Reference conditions for each of the selected nearshore fish metrics, determined from standardised historical and current seine net data collected from each region of the Swan Estuary (Lower Swan-Canning Estuary [LSCE], Canning Estuary/Lower Canning River [CELCR], Middle Swan Estuary [MSE] and Upper Swan Estuary [USE]) in each season. *n* = number of samples per region*season combination. + and - indicate positive and negative metric responses to degradation, respectively.

		Metric										
Region*season	*n*	*No species* (-)	*Prop trop spec* (-)	*No trop spec* (-)	*No trop gen* (+)	*Prop detr* (+)	*Prop benthic* (-)	*No benthic* (-)	*Prop est spawn* (-)	*No est spawn* (-)	*Prop P. olorum* (+)	*Tot no P. olorum* (+)
LSCE*summer	174	11	0.99	8	1	0	1.0	9	0.96	5	0	0
LSCE*autumn	156	13	0.99	8	1	0	1.0	9	0.83	5	0	0
LSCE*winter	173	8	1.0	6	0	0	1.0	6	0.79	4	0	0
LSCE*spring	179	11	0.98	7	1	0	1.0	8	0.76	5	0	0
CELCR*summer	66	14	0.99	9	1	0	1.0	9	1.0	9	0	0
CELCR*autumn	68	13	0.99	8	0	0	1.0	6	1.0	7	0	0
CELCR*winter	79	10	0.99	5	0	0	1.0	5	1.0	6	0	0
CELCR*spring	84	12	0.98	8	1	0	1.0	7	1.0	8	0	0

MSE*summer	119	14	0.96	8	1	0	1.0	9	1.0	9	0	0
MSE*autumn	123	14	1.0	9	0	0	1.0	9	1.0	8	0	0
MSE*winter	115	10	0.98	6	0	0	1.0	7	1.0	6	0	0
MSE*spring	144	13	0.93	8	1	0	1.0	9	1.0	8	0	0
USE*summer	108	10	0.98	6	1	0	0.98	7	1.0	8	0	0
USE*autumn	111	9	1.0	5	0	0	1.0	6	1.0	7	0	0
USE*winter	99	5	0.99	3	0	0	0.95	3	1.0	4	0	0
USE*spring	132	9	0.98	5	1	0	1.0	6	1.0	7	0	0

Table 5 Reference conditions for each of the selected offshore fish metrics, determined from comparable historical and current gill net data collected from each region of the Swan Estuary (Lower Swan-Canning Estuary [LSCE], Canning Estuary/Lower Canning River [CELCR], Middle Swan Estuary [MSE] and Upper Swan Estuary [USE]) in each season. *n* = number of samples per region*season combination. + and - indicate positive and negative metric responses to degradation, respectively.

		Metric						
Region*season	*n*	*No species* (-)	*Sh-div* (-)	*No trop spec* (-)	*No trop gen* (+)	*Prop detr* (+)	*Prop benthic* (-)	*Prop est spawn* (-)
LSCE*summer	11	6	1.51	4	0	0	1.0	1.0
LSCE*autumn	12	6	1.63	4	0	0	1.0	0.92
LSCE*winter	12	8	1.87	5	0	0	1.0	0.41
LSCE*spring	8	5	1.47	5	0	0	1.0	1.0
CELCR*summer	10	7	1.71	4	0	0.20	1.0	0.83
CELCR*autumn	8	8	1.69	4	0	0.36	1.0	0.72
CELCR*winter	10	4	1.36	3	0	0	1.0	1.0
CELCR*spring	8	9	1.71	4	0	0	0.96	1.0
MSE*summer	37	6	1.67	2	0	0.09	1.0	1.0
MSE*autumn	45	6	1.44	3	0	0.16	1.0	1.0
MSE*winter	42	5	1.44	2	0	0	1.0	1.0
MSE*spring	42	5	1.29	2	0	0.20	1.0	1.0
USE*summer	35	5	1.18	2	1	0	1.0	1.0

USE*autumn	39	5	1.55	3	0	0	1.0	1.0
USE*winter	39	4	1.18	1	0	0	1.0	1.0
USE*spring	37	4	1.27	1	1	0	1.0	1.0

List of Figures

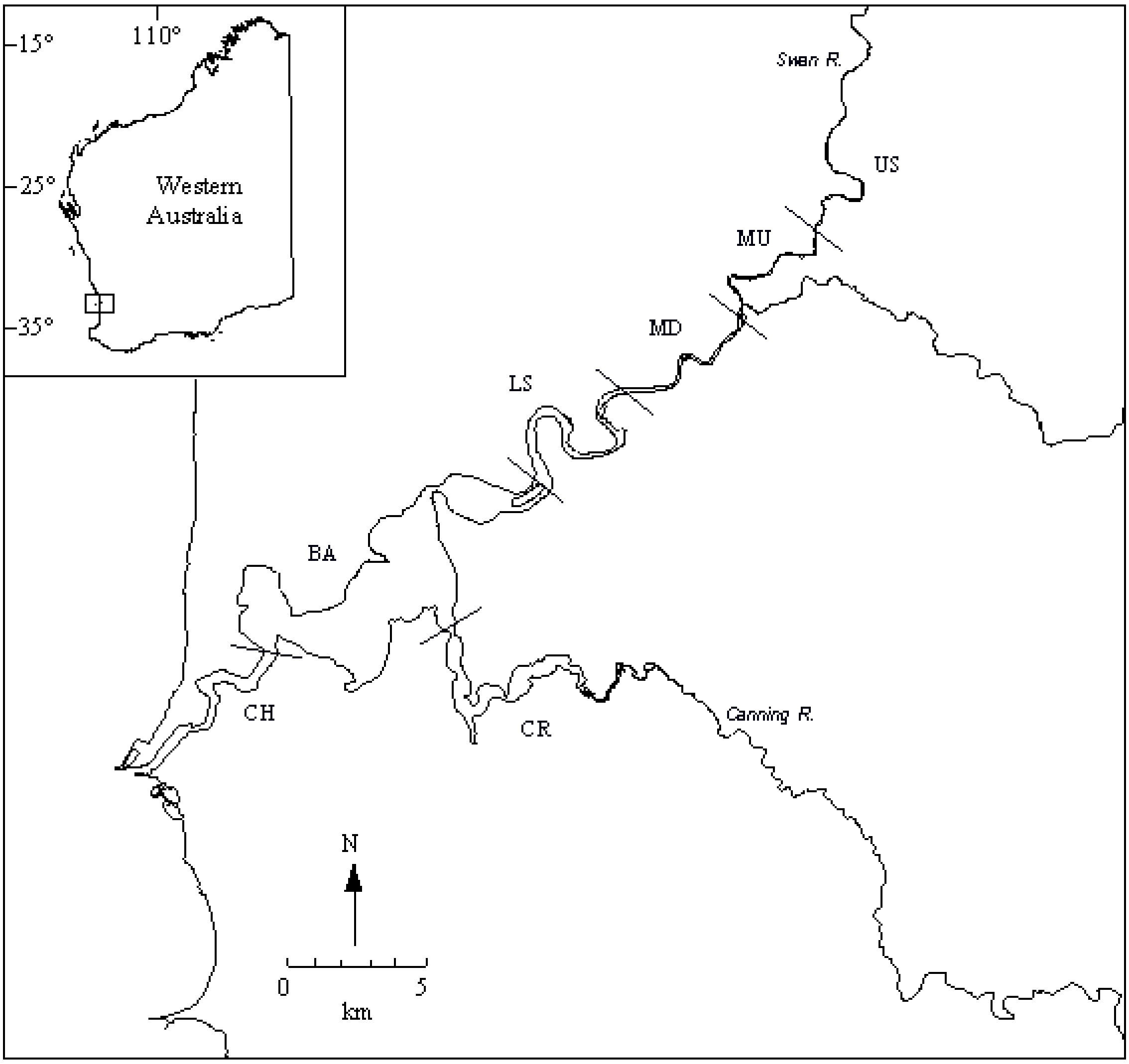

Figure 1 The Swan Estuary, Western Australia, and its location within Australia (inset), annotated to show regions of the estuary. CH = Channel, BA = Basin, CR = Canning River, LS = Lower Swan River, MD = Middle-Downstream Swan River, MU = Middle-Upstream Swan River, US = Upper Swan River.

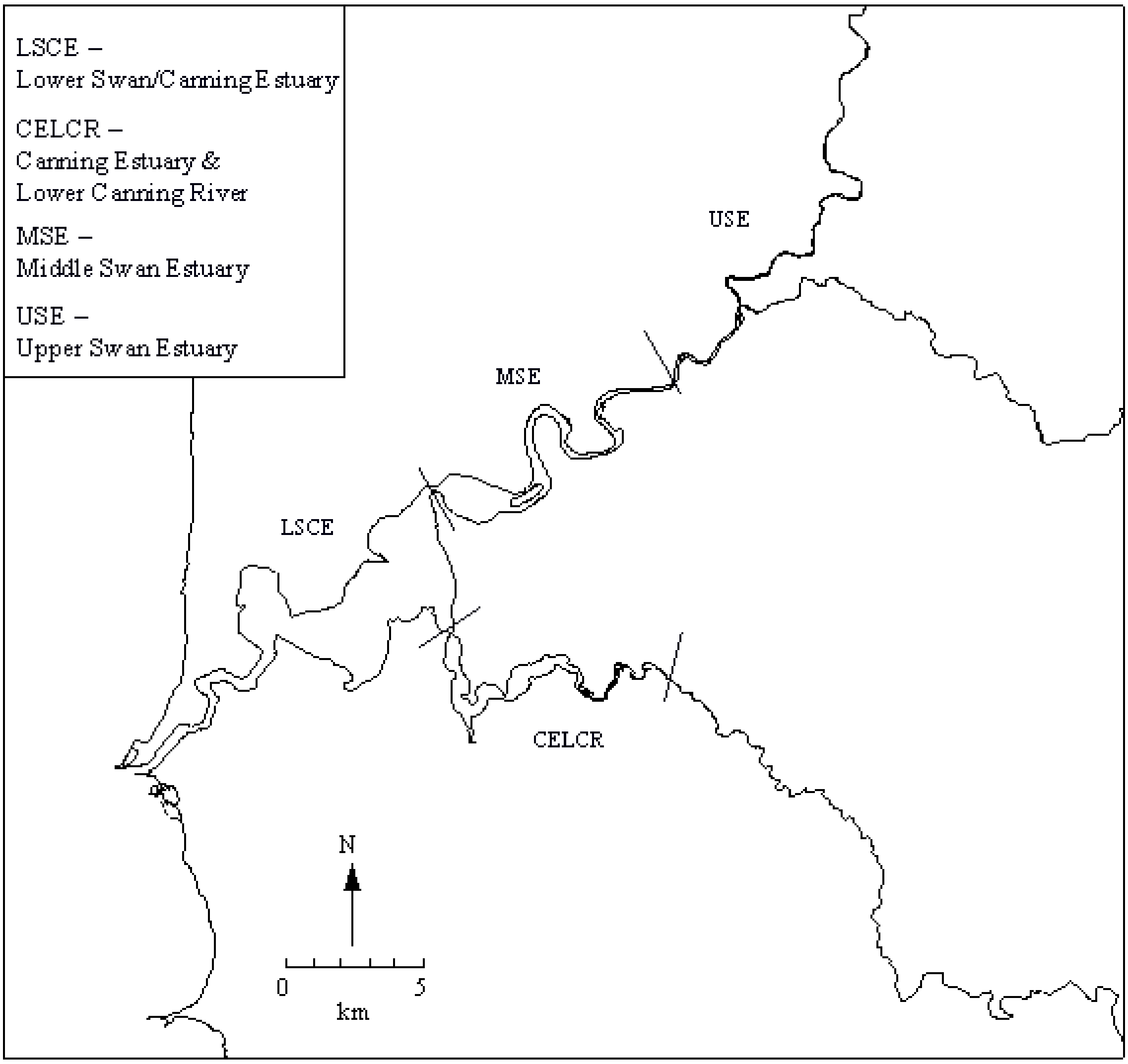

Figure 2 Regions (Ecological Management Zones) of the Swan Estuary for which specific reference conditions were established for each of the nearshore and offshore metrics.

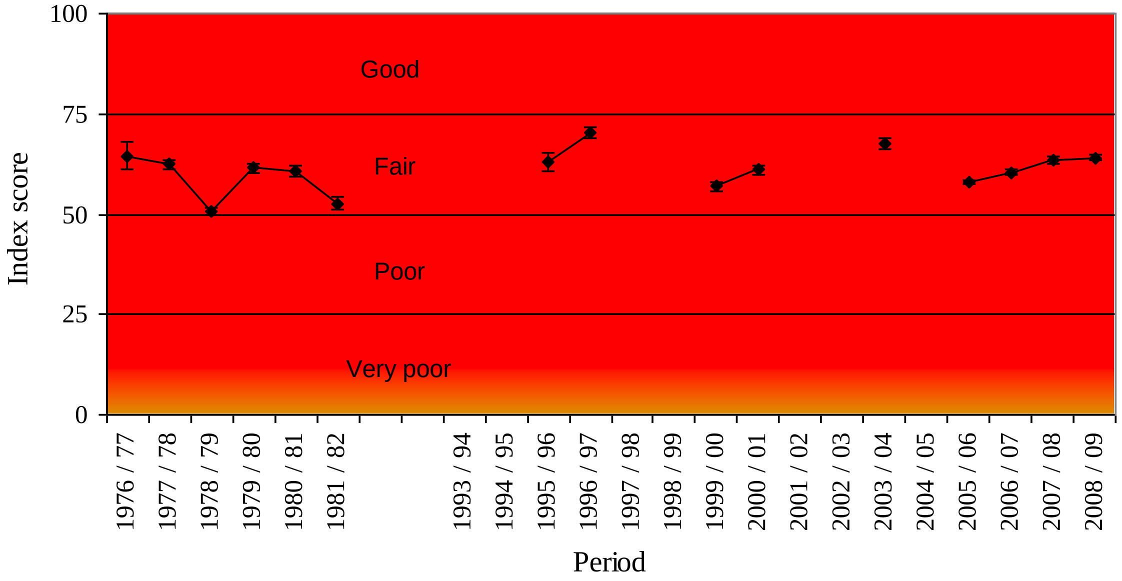

Figure 3 Mean (± SE) nearshore health index scores across all sites sampled throughout the Swan Estuary from 1976 to 2009.

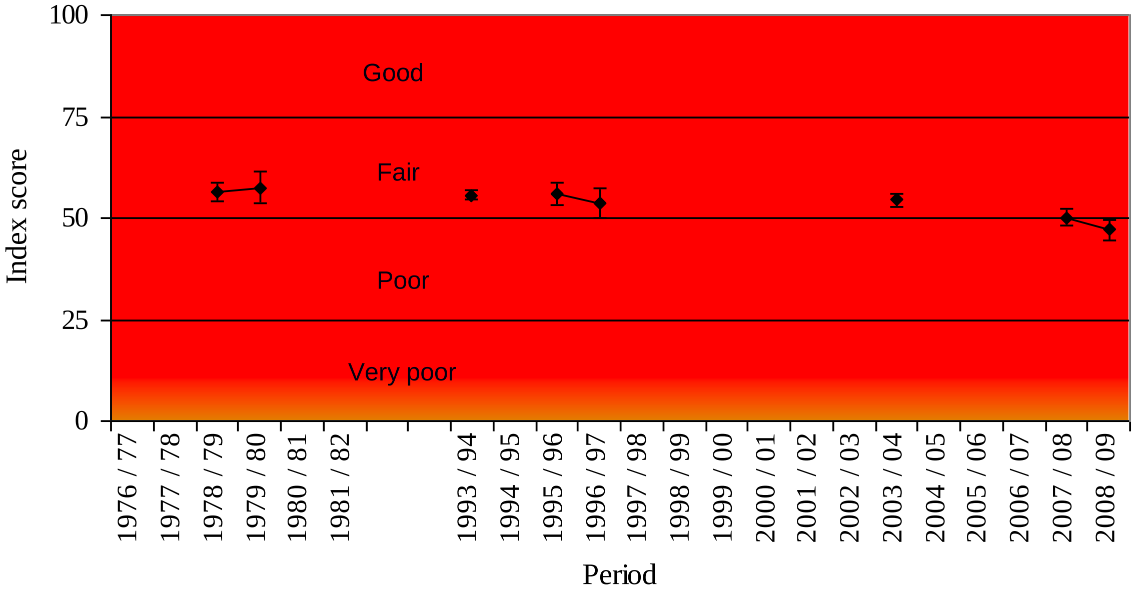

Figure 4 Mean (± SE) offshore health index scores across all sites sampled throughout the Swan Estuary from 1978 to 2009.

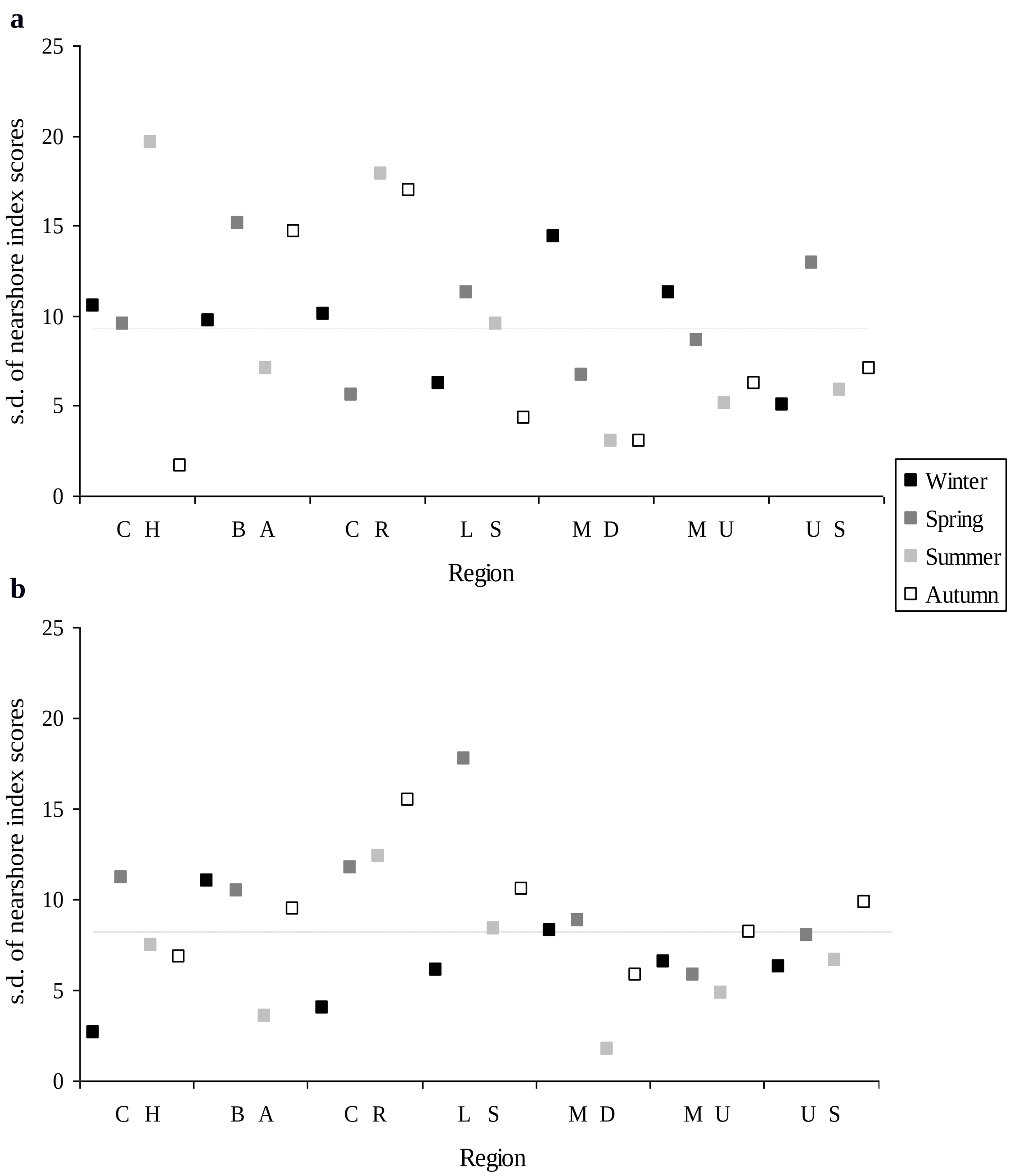

Figure 5 Plots of the standard deviation (s.d.) in nearshore health index scores among the three sites within each region of the Swan Estuary in each season during (a) 2007/08 and (b) 2008/09. See Fig. 1 for region codes. Dashed lines represent the average inter-site variability for each year, across all regions and seasons.

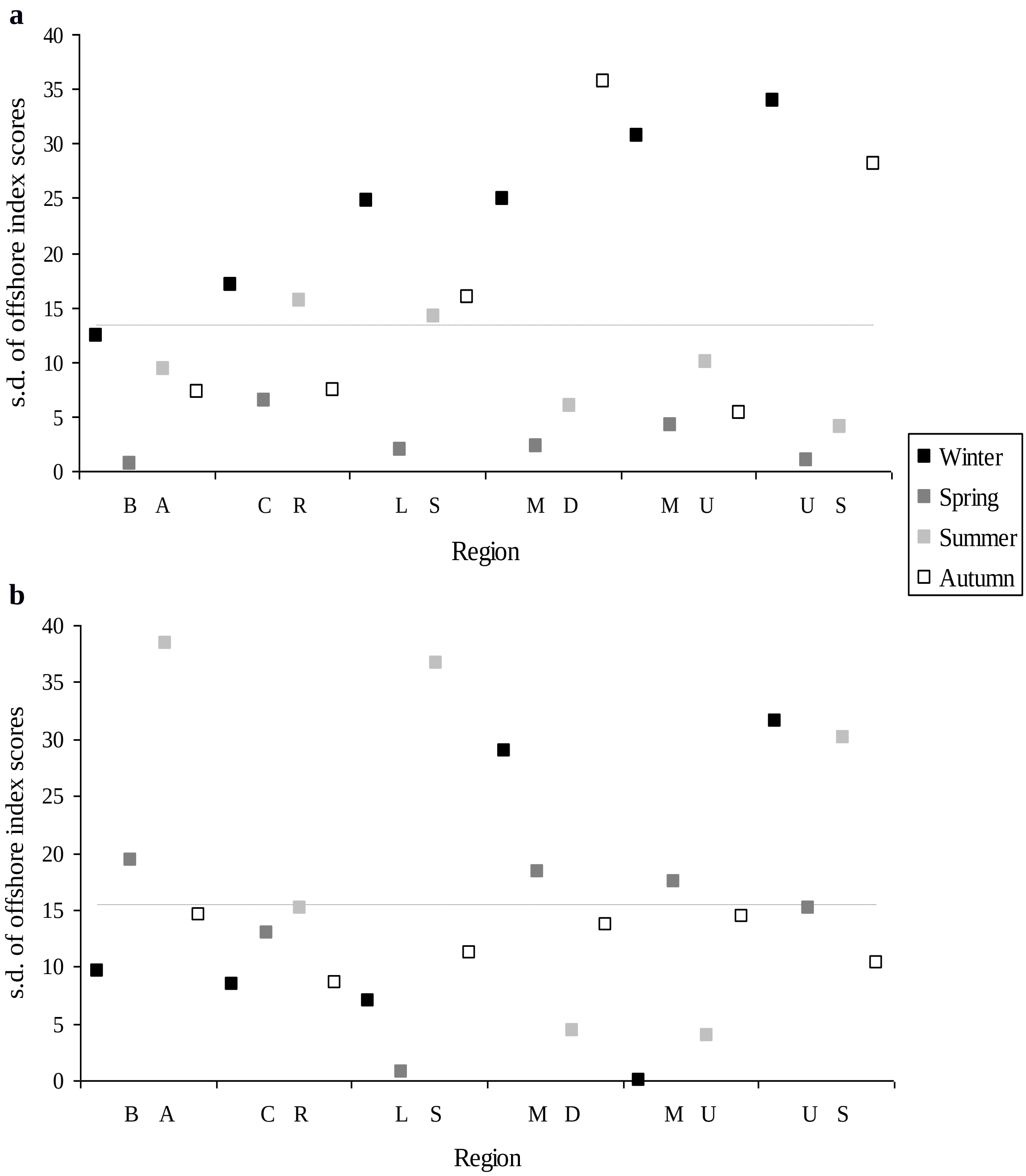

Figure 6 Plots of the standard deviation (s.d.) in offshore health index scores among the three sites within each region of the Swan Estuary, in each season during (a) 2007/08 and (b) 2008/09. See Fig. 1 for region codes. Dashed lines represent average inter-site variability for each year, across all regions and seasons.

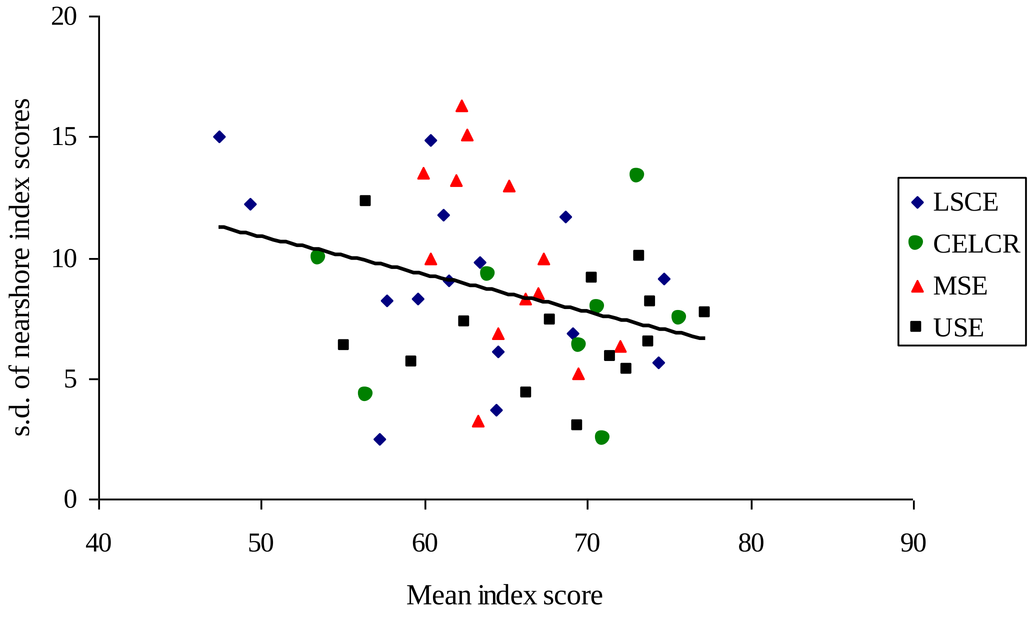

Figure 7 Mean *vs* standard deviation (s.d.) of nearshore health index scores among seasons at each of the sites sampled in 2007-2009. Sites are colour-coded for region of the estuary (see Fig. 2 for region codes). The solid line is a simple linear regression.

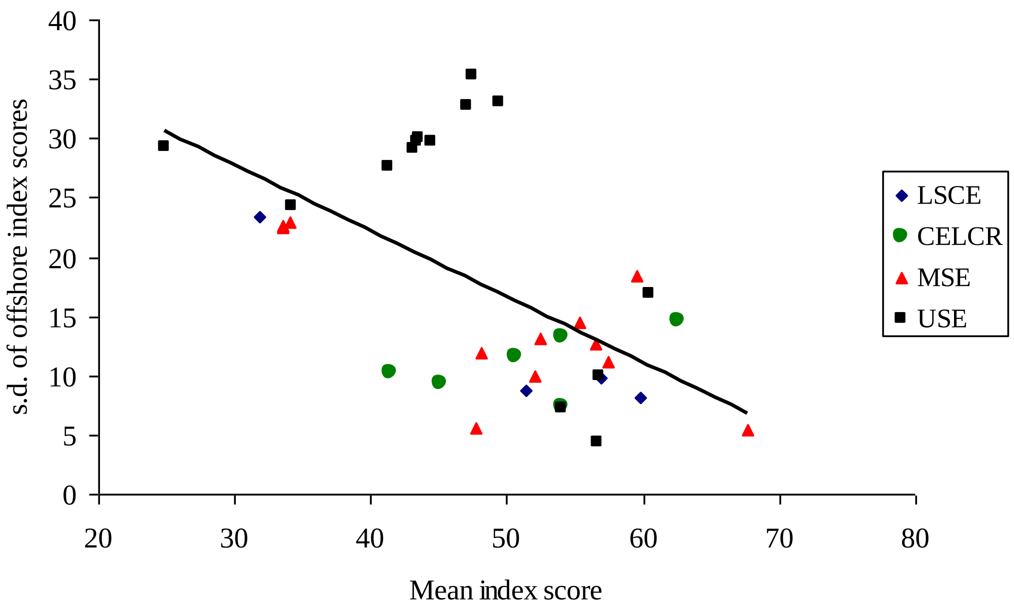

Figure 8 Mean *vs* standard deviation (s.d.) of offshore health index scores among seasons at each of the sites sampled in 2007-2009. Sites are colour-coded for region of the estuary (see Fig. 2 for region codes). The solid line is a simple linear regression.

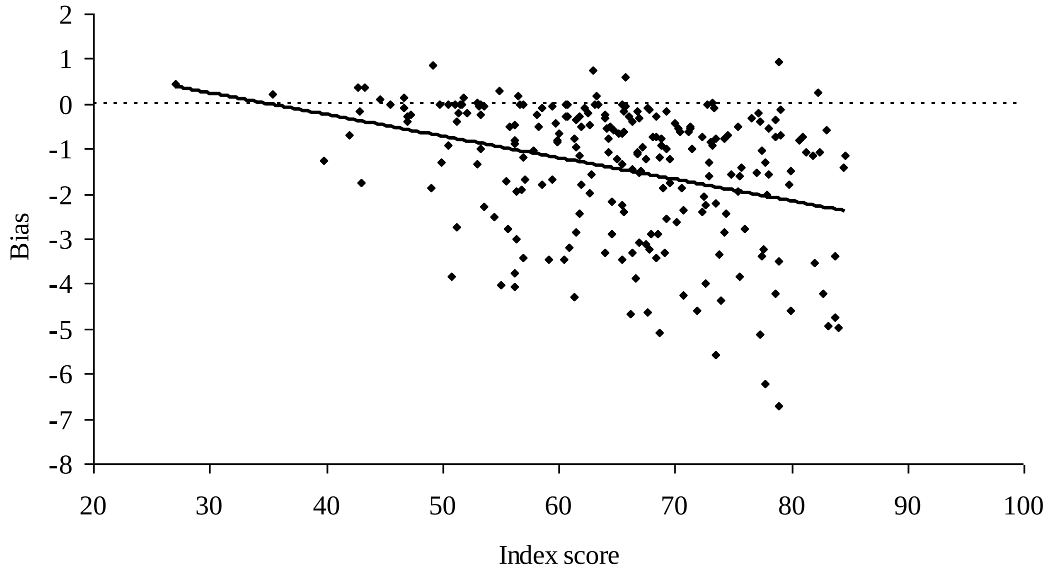

Figure 9 Bias (mean bootstrapped index score minus original index score) of the nearshore index scores from each site visit throughout the Swan Estuary in 2007-09. Dashed line represents zero bias expected if bootstrapped index scores matched original index scores. Solid line is the simple linear regression of bias as a function of original index score.

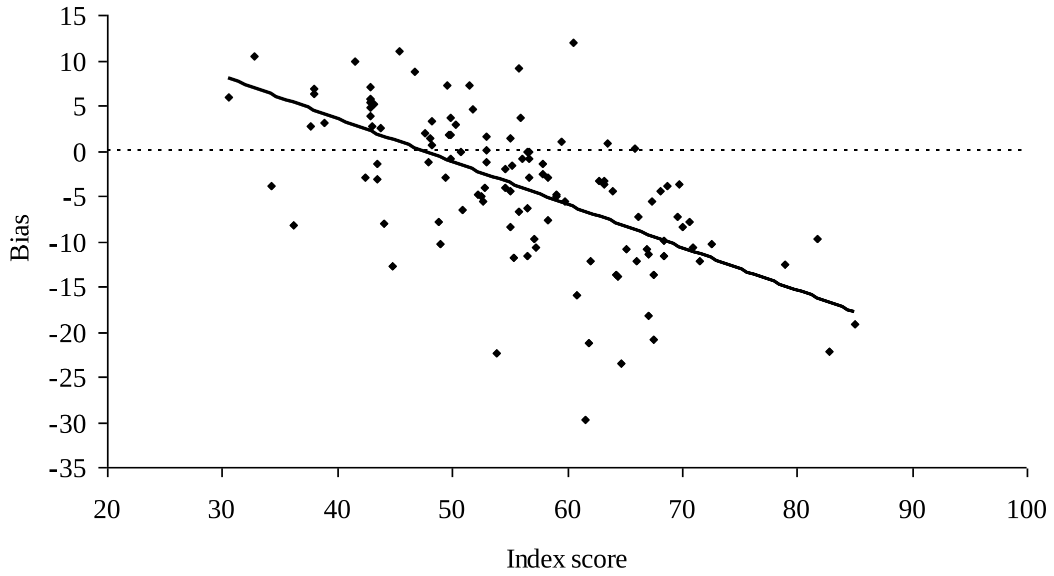

Figure 10 Bias (mean bootstrapped index score minus original index score) of the offshore index scores from each site visit throughout the Swan Estuary in 2007-09. Dashed line represents zero bias expected if bootstrapped index scores matched original index scores. Solid line is the simple linear regression of bias as a function of original index score.