Method for Analyzing Fish Assemblage Distribution with Application to Fishery Landings of Tropical Shallow Lake as Songkhla Lake, Thailand

Sarawuth Chesoh

Pattani Inland Fisheries Research and Development Center

Department of Fisheries, Ministry of Agriculture and Co-operatives, Pattani, 94160, Thailand

Tel: 66-73-439-123   E-mail: chesoh.s@hotmail.com

Chamnein Choonpradub (Corresponding author)

Department of Mathematics and Computer Science, Faculty of Science and Technology

Prince of Songkla University, Pattani Campus, 94000, Thailand

Tel: 66-73-312-179   E-mail: cchamnei@bunga.pn.psu.ac.th

Abstract

Fish community structure can provide potentially powerful tools for assessing aquatic environmental health. Monthly catch weights in Songkhla Lake were recorded over the period January 2003 to December 2006 inclusive for each of 127 species: 72 were marine vertebrate; 22 freshwater vertebrate; 21 marine invertebrate; 10 diadromous, and 2 catadromous. Linear regression model base on the log-transformed catch weights classified species and months, using these factors as multiplicative determinants based on principal components were performed, enabling assessment of clustering of species. The model has four such components which correspond to predominant seasonal time series patterns, giving an r-squared value of 87.7%. Purely seasonal patterns were identified for the first two components: estuarine and marine vertebrates showed considerable seasonal fluctuations but otherwise appeared to be steady over the four year period. Trends, mainly confined to the most recent year (2006) were identified for the third and fourth components: freshwater and estuarine fish had increasing catch weights, while the catch weights of marine invertebrates decreased. This model can provide practical lake information and reinforces that migratory fish species in tropical shallow lake need to be managed for sustaining their diversities.

Keywords: Catch pattern, Principal component regression, Fish clustering, Species diversity, Similarity index, Aquatic resource management

1. Introduction

A "shallow lake" is usually defined as a permanent standing body of water that is sufficiently shallow to allow light penetration to the bottom sediments adequate to potentially support photosynthesis of higher aquatic plants over the entire bottom (Wetzel, 2001). Fish assemblages are an important component of aquatic ecosystems of the lake basin and recognized as sensitive indicators of habitat disturbances, environmental deterioration, and overall ecosystem productivity (Gregory et al, 2009).

Songkhla Lake, the largest lagoonal water body of Thailand, covering 8,729 sq. km of the Lake Basin or 1,017 sq. km of main lake water body, is shallow (depth 1-2 m) and located on the lower east coast of the peninsular opening to the Gulf of Thailand between latitudes of 7° 10' to 7° 50' N and longitudes of l00° 05' to 100° 40' E (Figure 1). The Basin spans about 150 kilometers from north to south and about 65 kilometers from east to west. In addition, it has a multifunctional ecosystem ranging from tropical rain forest in upstream watershed area to a complex regime of water quality: freshwater, brackish and saline water with tidal and sea water intrusion influences from the Gulf of Thailand, runoff in monsoon seasons via twelve major rivers and various streams, and general drainage (Ratanachai and Kanchanasuwan, 2005). Tropical shallow lakes have high biodiversity and are also threatened globally from anthropogenic pressures and looming global climate change (Cairns and Lackey, 1992; Gopal, 2005; Enric et al, 2007). Similarity, Songkhla Lake, the largest lagoonal water body of Thailand, is one example of a tropical shallow lake in Southeastern Asia where facing critical euthophication and loss of fish population (Chesoh et al, 2008; Chesoh and Lim, 2008).

Monitoring of fish communities is advocated as an alternative to water quality monitoring for assessing ecosystem integrity. Actually, some fish species are ubiquitous in all habitats, mostly are migratory species: anadromous, catadromous, amphidromous, and oceanodromous, and so have specific habitats in Songkhla Lake (Choonhapran, 1996).

Several multivariate methods have been used to explain the assemblage structures and distribution patterns of fish including multivariate analysis of variance (Gauch, 1982; Jackson and Harvey, 1989; Ahmadi-Nedushan, et al, 2006).Unfortunately, ecological data are also some of the most complex, especially at large spatial and temporal scales concerning species composition and environmental factors, to analyze while the available resources often limit sample sizes. Statistically, principal components analysis (PCA) is an effective method of addressing the problems of large numbers of variables, multicollinearity and small sample sizes and PCA is also the approach most widely used to reduce the size of complex ecological data without losing information inherent in the data (Brazner and Beals, 1997, Vaughan and Ormerod, 2005, Chen el al, 2008).   In this study, we attempt to analyze fish assemblage patterns across time and across species in catch weights of fishery landings from Songkhla Lake over a four years period. We applied an interesting principal components regression approach, a regression-based method for successfully dealing with multicollinearity and giving results in estimation and prediction better than ordinary least squares (Fekedulegn et al, 2002), and revealed different temporal patterns across species. A form of ordination of regression coefficients was used to break the fish species into each distinct category, again using interesting methodology that is not widely presented in this discipline.

The advantage of this study is that it provides a statistically valid model for measurement of ordering of species. It can then be used to provide essential information for ecosystem-based approaches to fisheries management in tropical shallow lakes as Songkhla Lake of Thailand.

2. Materials and Methods

Study area and data source

Songkhla Lake is shallow (depth 1-2 m) and located on the lower east coast of the peninsular opening to the Gulf of Thailand between latitudes of 7° 10' to 7° 50' N and longitudes of l00° 05' to 100° 40' E (Figure 1). The Basin spans about 150 kilometers from north to south and about 65 kilometers from east to west. The water regime is   complex, with tidal and sea water intrusion influences from the Gulf of Thailand, runoff in monsoon seasons via twelve major rivers and various streams, and general drainage.

There are ten major fish catch landing sites around the entire Songkhla Lake (Figure 1): Khu Tao (KT), Kuan Nieng (KN), Pak Pa Yoon (PY), Jong Ke (JK), Lampam (LP), Thale Noi (TN), Ra Nod (RN),   Ko Yai (KY), Khu Kud (KK) and Hua Khao Daeng (HD). These sites were selected for data collection from January 2003 to September 2005 by the National Institute of Coastal Aquaculture (NICA) of the Department of Fisheries of Thailand, and thereafter to December 2006 by the current authors. Data include daily records of species, weight of the catch; gear types used, and catch value. Fish samples were classified following scientific taxonomic systematics. In addition, six categories were defined in terms of biological and habitat characteristics: vertebrate or invertebrate, and freshwater, estuarine and marine habitat, following Choonhapran, 1996.

[Figure 1 about here]

Statistical Analysis

Various statistical methods are available for clustering aquatic and marine organisms according their patterns of variation in space and time (Hawkins et al, 2000; Joy and Death, 2000; Fre´dou et al, 2006). Clarke and Warwick (1994) have outlined many of these methods in detail. They include data transformation using square roots, fourth roots or logarithms to remove skewness, principal components analysis of covariance matrices, and ordination procedures based on Bray-Curtis similarity indexes giving multidimensional scaling (MDS) plots used to cluster taxa in space and time.

In this study, seasonal comparisons of fish abundance in terms of catch weight were transformed by taking natural logarithms to address statistical analysis requirements. Since the total catch weight (*wt*) for a given species in a specific month can be zero, we used the transformation

$$y = \ln(wt + 1).$$ \hfill (1)

After preliminary analysis of average monthly catch weights and corresponding monetary values for each species over the four-year period, and changes in the annual total catch weights for each species from one year to the next, we then fitted a linear regression model to the log-transformed monthly catches. The model used is an extension of two-way analysis of variance (ANOVA) incorporating additional terms based on a PCA (Good, 1969; Theil, 1983; McNeil and Tukey, 1973; Booth et al, 2002). Briefly, PCA is a multivariate technique for examining the relationship between several quantitative variables. Fisheries data are commonly subject to atypical errors; especially the large number of variables measured, highly skewed and zero-inflated, resulting in outliers when modeling. A PCA can be used to reduce the size of a large set of the original variables with a few PCs without losing information inherent in the data (Cooley and Lohnes, 1971; Jackson and Chen, 2003; Chen et al, 2008).

The two-way ANOVA method is the simplest regression model that allows for differences between species (*s*) and month (*t*), and is expressed as the additive combination

$$y_{st} = \mu_s + \nu_t, \tag{2}$$

This model has $S + T$ parameters, where S and T denote the numbers of fish species groups and observation months, respectively. It assumes that the distribution of a species group over the observation period is the same for all species groups, only differing in level through the parameter $_s$. Similarly the model assumes that the temporal pattern $_t$ is the same for all species groups.

PCA was performed on the correlation matrix from the complete set on 127 fish species and 48 months (2003 to 2006) of catch weights data record. Equation (2) can be generalized to overcome these limitations by defining predictor variables as principal components $\beta_t^{(k)}$ defined as the eigenvectors of the covariance matrix of the data, ordered by decreasing size of their corresponding eigenvalues (see, for example, Johnson and Wichern, 1998), namely,

$$y_{st} = \mu_s + \sum_{k=1}^{m} \alpha_s^{(k)} \beta_t^{(k)} \tag{3}$$

In this formulation the data matrix has successive months as its T column variables and species as its S rows and we assume $S > T$. Each eigenvector is scaled to have sum of squares equal to 1 and each pair of eigenvectors has sum of products equal to 0. The number of predictor variables selected for inclusion in the model will depend on biological considerations and the amount of total sample variance explained. The model (eqn 3) has taxa-specific parameters $_s$ encapsulating the variation in catch weights between species, and m sets of coefficients $\alpha_s^{(k)}$, $k = 1,2, \ldots, m$, denoting the extent to which the taxa have each of the m specific time-changing patterns. A detailed analysis of the goodness-of-fit that highlights individual anomalies involves graphing residuals against corresponding quantiles from the standardized normal distribution.

An advantage of this method over the more conventional approaches described by Clarke and Warwick (1997) is that it routinely provides standard errors for the taxa-specific parameters $\alpha_s^{(k)}$, $k = 1,2, \ldots, m$, which in turn provide a valid statistical basis for pairwise comparisons of species based on chi-squared tests on Euclidean distances between their locations in the corresponding m-dimensional space (Anderson and Millar, 2004; Anderson, 2005; Legendre, 2005; Legendre et al, 2005).

We used the R statistical system (Venables and Smith, 2004) for statistical model fitting, assessing the goodness-of-fit, and plotting data, fitted models, parameters and confidence intervals.

3. Results

General catch information

During the 4 year study period from January 2003 to December 2006, the mean annual catch in Songkhla Lake was 2,499.9 tonnes (range 2,388.2 – 2,643.0). These fish were caught by three major types of fishing gear: set-bag net (64.7% of catch weight), followed by traps (21.8%) and gill nets (13.5%). The most productive fishing ground was the Lower Lake and this accounted for 60% of the total annual catch, comprising the major groups of high economic importance among brackish and marine fish species. The second most productive fishing ground was the Middle Lake (22.8%), dominated by brackish and euryhaline fish, and followed by the Upper Lake (17.2%) where the most abundant fish in the catch were freshwater fish species. A total of 127 aquatic animal species belonging to 68 families were caught.

Figure 2 shows normal quantile plots of the 6,096 (48 months by 127 species) monthly catch weights, before and after taking the transformation (1). The distribution of the raw weights is highly skewed (skewness coefficient 18.6) with the bulk of the catch weights less than 10,000 kg. There were only two monthly catches for any species exceeding 50,000 kg, namely 158,230 kg for greasy back shrimp in March 2003 and 149,240 kg for short-nose pony fish in March 2006. Even when these two outliers are omitted the skewness is still high, though reduced to 5.7. In contrast, the transformed catch weights are well approximated by a normal distribution.

[Figure 2 about here]

Figure 3 shows scatter plots of the average catch weights per month for each of the 127 species over a four year period of the study (horizontal axis) and corresponding monetary values (vertical axis). The data are plotted linearly in the left panel and using log scales on each axis in the right panel. The alphabetic symbols denote species with noteworthy characteristics (see Table 1 for codes).

[Figure 3 about here]

The monetary values (in 1,000s of Baht per tonne) are computed simply by multiplying the average catch weights by the corresponding price per unit in weight. The four leading species (Broadhead anchovy, Sumatran silverside, Chacunda gizzard shad, and Black tiger shrimp) accounted for 21.4% of the total catch weight, contributing 6.35, 5.32, 5.28 and 4.46 percent, respectively. In monetary value, the four most valuable species (Black Tiger Shrimp, Chacunda gizzard shad, Sumatran silverside, and Greasy Grouper) accounted for 23.7% of the total value of the catch,

contributing 10.67, 5.05, 4.24, and 3.70 percent, respectively. The fish with highest value (the Greasy Grouper; 250 baht per kilogram) made a relatively small contribution to the catch weight (1.95 tonnes per month). In contrast, the leading species (Broadhead anchovy; 30 baht per kilogram and used mainly for duck food) made a lesser value contribution.

Annual changes in catch weights from 2003-2004, 2004-2005 and 2005-06 for each species are shown in Figure 4. Most estuarine (green colour) and marine (magenta) invertebrates exhibited overall serious decline from 8.3%, 0.4% and 13.4%, followed by freshwater fish (blue and one brown colour) from 9.2%, 23.9% and 1.4%, respectively, while estuarine (red) and marine (grey) vertebrate exhibited quite steady change in study period from 16.7%, 3.1% and 15.4%.

[Figure 4 about here]

Temporal patterns of catch species diversity

The catch weights of 127 fish species in 48-monthly collections from 2003 to 2006 were modeled using Equation (3).

The goodness of fit of the model shows that $m = 4$ components fit the data reasonably well, with r^2 equal to 0.877, although the residuals plot indicates some departure from the statistical normality assumption. The r^2 value for the simple additive model based on Equation (2) was 0.722, whereas the values for model (3) were 0.776, 0.829 and 0.859 for $m = 1, 2$ and 3, respectively. Because the four components each have sum of squares equal to 1, the standard errors for the $\alpha_s^{(k)}$ coefficients are all the same, having estimated value 0.489. (Note that the   . coefficients reflect the overall catch weights without regard to month-by-month changes, and thus depend largely on the biomass of the various species in the Lake.)

From the distribution of eigenvector plots of each of four components $\beta_t^{(k)}$, $k = 1, 2, 3$ and 4, all series display interpretable temporal patterns. The first component shows a similar seasonal pattern for the whole period, with a spike occurring in March of each year. As shown in Figure 5, this pattern is characterized by species like Russell's snapper (*Lutjanus russellii*) Red tilapia; alien hybrid fish (*Oreochromis niloticus   X   O. Mossambicus*), and Horse face roach (*Acantopsis choirorhynchos*), as shown in Figure 6.

[Figure 5 about here]

[Figure 6 about here]

The second component also shows a regular seasonal pattern with peaks occurring in February and declining to a minimum in December of each year after dipping in March. There is no species that solely exhibits this pattern, but it is clearly present in Spanish mackerel (*Scomberomorus commerson*), Dusky Jack (*Caranx sexfasciatus*) and Indo-pacific mackerel (*Scomberomorus guttatus*), as Figure 7 shows. However, these three species have maximum catch weights in March, indicating sufficient presence of the first component to offset the second component's March dip.

Figure 8 also shows three species that are characterized by third model component pattern: three-spot gourami (*Trichogaster trichopterus*), Croaking gourami (*Trichopsis vittata*), and Snakeskin gourami (*Trichogaster pectoralis*) with peaks occurring in November and declining to a minimum in March and April of each year.

Figure 9 shows three species that to some extent follow the pattern of the fourth component are characterized as species black tiger prawn (*Penaeus monodon*), black lancer catfish (*Mystus cavasius*), and shortnose ponyfish (*Leiognathus brevirostris*) with the slightly increasing trend to the recent year. In contrast to the regular seasonal cycles exhibited by the first two principal components, the third and fourth components are less regular and show an upward trend that rises more sharply in the most recent year of the study period.

[Figure 7 about here]

[Figure 8 about here]

[Figure 9 about here]

Figure 10 shows scatter plots of the component coefficients $\alpha_s^{(k)}$, $k = 14$, with the first two components plotted against each other on the left and the third and fourth on the right. In each plot pairs of points are joined if the distance between them in their four-dimensional space is not statistically significant at the 5% level. These p-values are the probabilities that a chi-squared statistic with 4 degrees of freedom exceeds $D/(2\sigma^2)$, where D is the Euclidean distance between the two points and σ is the common standard error of the $\alpha_s^{(k)}$ coefficients (0.489).

[Figure 10 about here]

Two distinct clusters are clearly separated. The smaller cluster contains all freshwater species of vertebrates (blue symbols) and the single freshwater invertebrate (the giant freshwater prawn, brown, labeled g), whereas the second group contains all marine vertebrates (grey) and invertebrates (green). The estuarine species, red for invertebrates and magenta for vertebrates, appear in both clusters.

The graph also shows a small disconnected cluster of marine vertebrates labeled by the symbols T (Spanish mackerel), U (dusky Jack) and e (Indo-Pacific mackerel), and singleton disconnected species labeled by the symbols V (greasy back shrimp), N (black lancer catfish), X (sand goby), a (spotted green pufferfish), b (cuttlefish), c (largescale tonguesole), d (spotted codlet) and f (lined silver grunt). The other points labeled with lower-case letters from g to z comprise all the remaining invertebrates, identified individually in Table 1, and includes all species not connected to other species of the same type within the same cluster.

[Table 1 about here]

Note that 12 of the 18 marine invertebrates are located within the smaller of the two clusters together with all the freshwater fish and two of the four estuarine invertebrates, whereas the other six are located in the larger cluster of estuarine and marine vertebrates.

4. Discussion

Our statistical model involves four components and gave an r-squared value of 87.7%, with the first four components, 67.3 % of the total variance could be explained. Although the model based on equation (2) had the correlation matrix from a data set of 127 species and 48 months and equation (3) had *m* equal 4 whereas *S* equal 127 that given a large number of parameters. Jong and Kotz (1999) reported the results of merging the concepts of PCA and multivariate regression showing the equivalence of optimization criterion involved in each one of them and also presented the scale invariance property of PCA derived by regression approach. This property allows us to rescale PCA without changing their capabilities. As previous mentions, PCA is a well known technique the aim of which is to synthesize huge amounts of numerical data by means of a low number of unobserved variables, called components while as least-squares techniques, commonly known, are not robust in the sense that outlying measurements can arbitrarily skew the solution from the desired solution (Hampel et al, 1986; Jolliffe, 1986). PCA is also of interest because they describe the predominant temporal (seasonal and trend) patterns present in the data. Even where trends are relatively small compared to seasonal patterns, they can highlight features of practical importance that might reflect lack of sustainability the fish catch or environmental degradation in the Lake. Throughout the study, the model was effective in clearly separating four distinctive fish community clusters. The first two components show purely seasonal patterns; the first pattern shows a spike occurring in March and the second has a peak in February with a gradual decline to December followed by a sharper increase. The third and fourth patterns show less pronounced seasonal effects with a trend increasing in the most recent year (2006).

Freshwater fish, together with black tiger shrimp, giant freshwater prawn, white seabass and mullets showed increasing catch weights, while other Penaeid shrimp (greasy back shrimp, green tiger prawn and small white shrimp), Mantis shrimp, Hamilton's thryssa (*Thryssa dussumieri*), and Chacunda gizzard shad decreased. In fact, the overall fishing effort in the Lake was fairly stable in time because both fishing gear and number of fishermen did not change substantially (NICA, 2007). Therefore, the increasing trend of black tiger prawn, giant freshwater prawn, white seabass and freshwater catfish might be due to increased seeding, stock enhancement and fishery rehabilitation projects by governmental agencies (NICA, 2007). However, increasing catch quantities may signal over-fishing and sizable quantities of by-catch fish caught unintentionally. Catch weights of large tooth flounder (*Pseudorhombus arsius*), bigeyed sand goby (*Gnatholepis alliurus)*, short-nosed pony fish (*Leiognathus brevirostris*), naked-head glassy perchlet (*Ambassis gymnocephalus*) and starry triggerfish (*Abalistes stellaris*) increased whereas most estuarine and marine invertebrates and some high value freshwater fish had overall declines.

Plots of regression coefficients of principal components show two distinct clusters clearly separated between freshwater and saltwater fish. Some estuarine species (invertebrate and vertebrate) appear in both clusters because diadromous and euryhaline fish can adapt to a wide range of salinities and migrate between fresh and salt water (McDowall, 1995; Musick et al, 2001; Welcomme et al. 2006). For example, different life stages of Penaeid shrimp have distinct salinity preferences despite being marine invertebrate (Dall et al, 1990). Wirth et al (2004) reported that Penaeid shrimp can be raised in low salinity (1.56 ppt) geothermal water at inland sites without adverse effect on growth and survival. Furthermore, these estuarine and marine invertebrate are usually found in shallow, semi-enclosed estuarine bays in the southern Gulf of Thailand (Hajisamae et al, 2006), and generally spawn and spend much of their adult life in saltwater or offshore, but enter the Lake seasonally (Yanez-Arancibia et al, 1994; Hajisamae et al, 2003; Khongchai et al, 2003).

The croaking gourami (*Trichopsis vittata*) is very sensitive to salinity change from freshwater (Liengpornpan et al, 2004), and appears at the bottom left of Figure 7 (labeled Z) within a sub-cluster of marine fish. The high salinity offshore marine vertebrates such as Spanish mackerel (*Scomberomorus commerson*), Dusky Jack (*Caranx sexfasciatus*), and Indopacific mackerel (*Scomberomorus guttatus*) are found from the edge of the continental shelf to shallow coastal waters (Froese and Pauly, 2004), and appear at the top right of Figure 7 (labeled T, U and e, respectively). The spotted green pufferfish (a), spotted codlet (d), lined silver grunt (f), Greasy back shrimp (V) and Sand goby (X) appear as singletons disconnected to other major marine vertebrates, and have a less regular catches with substantial seasonal variation in monthly catch weights.

The salinity in various locations in the Lake also depends on the season, dropping substantially during the heavy monsoon that usually occurs from October to December (Chesoh and Lim, 2008). These seasonal patterns affect fish distribution and food web structure from different habitats (Winemiller and Jepsen, 1998; Thompson and Townsend, 1999; Gibson et al, 2000), and encourage major events in the life cycle of each species that take advantage of increased productivity. Alien fish species, namely Mozambique tilapia, hybrid red tilapia and African walking catfish, are increasingly found in the Lake and might destroy native species or alter the gene pool (Balon and Bruton, 1986; Salonen and Mutenia, 2007).

This study reflects the broad band of fish assemblage distribution according to distance from the Lake's junction with the Gulf of Thailand from saline to brackish water to freshwater, with some species confined to specific salinity bands, some euryhaline marine invertebrate species preferring to feed in low salinity biotopes, and others dwelling everywhere. Generally, the lake fishing is based on trapping the fish on their passage from their feeding ground to spawning and nursery grounds (Katselis et al, 2003). This seasonal pattern reinforces that these migratory fish species in tropical shallow lake need to be managed for sustaining their diversities.

Various manuals commonly used multivariate analysis package in ecology, in summarizing the methods that are currently used for analyzing fish community data. However, these methods are most commonly used for abundance as measured using counts of fish, presence or absences, or semi-quantitative measures, rather than monthly catch weights. Our model presents a valid statistical basis for pairwise comparisons of species using chi-squared tests on Euclidean distances in the ordination space, regression coefficients was used to break the fish species into distinct categories. However, the study does not take account of environmental factors and types of fishing gear used. Generally, multivariate regression models are performed by applying criterion of minimizing residual sum of square without employing the concepts of eigenvectors and eigenvalues. PCR purpose is not to obtain the best of the linear transformation matrix but rather to solve the problems of multicollinearity among predictor variables (Jong and Kotz, 1999). Therefore, application of linear regression approach to principal component analysis in fishery data will be widespread and increasing. We confirm that the robust techniques like those proposed will prove useful as linear models are used to represent more realistic data sets.

5. Acknowledgements

References

Anderson, M. J. & Millar, R. B. (2004). Spatial variation and effects of habitat on temperate reef fish assemblages in northeastern New Zealand. *Journal of Experimental Marine Biology and Ecology*, 305, 191-221.

Anderson, M. J. (2005). Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics*, 62, 245-253.

Balon, E. K. & Bruton, M. N. (1986). Introduction of alien species of why scientific advice is not heeded. *Environmental Biology of Fishes*, 16, 225-230.

Booth, H., Maindonard, J. & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56, 352-336.

Brazner, J. C. & Beals, E.W. (1997). Patterns in fish assemblages from coastal wetland and beach habitats in Green Bay, Lake Michigan: a multivariate analysis of biotic and abiotic forcing factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 54, 1743-1759.

Cairns, M. A. & Lackey, R. T. (1992). Biodiversity and management of natural resources: the issues. *Fisheries*, 17, 6-10.

Chen, Y., Chen, X. & Xu, L. (2008). Developing a size indicator for fish populations. *Scientia Marina*, 72, 221-229.

Chesoh, S. & Lim, A. (2008). Forecasting fish catches in the Songkhla Lake basin. *ScienceAsia*, 34, 335-340.

Chesoh, S., Lim, A. & Tongkumchum, P. (2008). Model predicting Chlorophyll-a concentration as an index of potential occurrence of eutrophication in Songkhla Lake, Thailand. In: Sengupta, M. & Dalwani, R. (Eds), pp 834-839. In Taal 2007: The 12th World Lake Conference, 28 October - 2 November 2007, Jaipur, India.

Choonhapran, A. (1996). Study on Fisheries Resources and Population Changes in Songkhla Lake: Case study on 3 Fishing Gears. Technical paper 18/1996, National Institute of Coastal Aquaculture (NICA), Songkhla, Thailand, pp 1-32.

Clarke, K. R. & Warwick, R. M. (1994). Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. Plymouth: PRIMERE Ltd, pp 1-144.

Cooley, W. W. & Lohnes, P. R. (1971). *Multivariate Data Analysis*. New York: John Wiley & Sons.

Dall, W., Hill, B. J., Rothlisberg, P. C. & Staples, D. J. (1990). Biology of the Penaeidae. In: Blaxter JHS, Southward AJ (Eds), Advances in Marine Biology, vol 27. London, UK: Academic Press, pp 1-489.

Enric, S., Knowlton, N. (Lead Authors) & Duffy, J. D. (Topic Editor). (2007). "Global marine biodiversity trends" In: encyclopedia of Earth. Eds. Cutler J. Cleveland (Washington, DC: Environmental Information Coalition, National Council for Science and the Environment). [Online] Available: http://www.eoearth.org/article/Global_marine_biodiversity_trends (March 6, 2008)

Frédou, T., Ferreira, B. P. & Letourneur, Y. (2006). A univariate and multivariate study of reef fisheries off Northeastern Brazil. *Journal of Marine Science*, 63, 883-896.

Fekedulegn, B. D., Colbert, J. J., Hicks, R.. R. Jr. & Schuckers, M. E. (2002). Coping with multicollinearity: an example on application of principal components regression in dendroecology. [Online] Available: http://www.fs.fed.us/ne (February 13, 2009)

Froese, R. & Pauly, D. (eds). (2004). FishBase. World Wide Web electronic publication version (10/2004). [Online] Available: http://www.fishbase.org (January 23, 2008)

Gauch, H. G., Jr. 1982. *Multivariate Analysis in Community Structure*. Cambridge University Press, Cambridge.

Gibson, G. R., Bowman, M.L., Gerritsen, J. & Snyder, B. D. (2000). Estuarine and Coastal Marine Waters: Bioassessment and Biocriteria Technical Guidance. EPA 822-B-00-024. Office of Water, Washington, DC: US Environmental Protection Agency.

Good, I. J. (1969). Some Applications of the Singular decomposition of a matrix. *Technometrics*, 11, 823-831.

Gopal, B. (2005). Does inland aquatic biodiversity have a future in Asian developing countries?. *Hydrobiologia*, 542, 69-75.

Gregory, S., Wildman, R., Ashkenas, L., Wildman, K. & Haggerty, P. (2009). Fish Assemblages. [Online] Available : http://www.fsl.orst.edu/pnwerc/wrb/Atlas_web_compressed/4.Biotic_Systems/4d.fish_assmbl_web.pdf (February 13, 2009)

Hajisamae, S., Choua, L. M. & Ibrahimb, S. (2003). Feeding habits and trophic organization of the fish community in shallow waters of an impacted tropical habitat. *Estuarine, Coastal and Shelf Science*, 58, 89-98.

Hajisamae, S., Yeesin, P. & Chaimongkol, S. (2006). Habitat utilization by fishes in a shallow, semi-enclosed estuarine bay in southern Gulf of Thailand. *Estuarine, Coastal and Shelf Science*, 68, 647-655.

Hampel, F., Ronchetti, E., Rousseeuw, P. & Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley & Sons.

Hawkins, C. P., Norris, R. H., Hogue, J. N. & Feminella, J. W. (2000). Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications*, 10, 1456-1477.

Hynes, H. B. N. (1970). The Ecology of Running Waters. Liverpool, Great Britain: Liverpool University Press, pp 1-555.

ILEC. (2005). Managing Lakes and their Basins for Sustainable Use: A Report for Lake Basin Managers and Stakeholders. Kusatsu, Japan: International Lake Environment Committee Foundation, pp 1-146.

Jackson, D. A. & Chen, Y. (2003). Robust principal component analysis of ecological data. *Environmetrics*, 14, 1-11.

Jackson, D.A. & Harvey, H. H. (1989). Biogeographic associations in fish assemblages: local versus regional processes. *Ecology*, 70, 1472-1484.

Johnson, R.A. & Wichern, D.W. (1998). Applied multivariate statistical analysis. (4th ed.). New Jersey: Prentice Hall, pp 1-816.

Jolliffe, I. T. (1986). *Principal Component Analysis*. New York: Springer-Verlag.

Jong, J. & Kotz, S. (1999). On a Relation between Principal Components and Regression Analysis. *The American Statistician*, 53, 349-351.

Joy, M. K. & Death, R. G. (2000). Development and application of a predictive model of riverine fish community assemblages in the Taranaki region of the North Island, New Zealand. *New Zealand Journal of Marine Freshwater Research*, 34, 241-252.

Katselis, G., Koutsikopoulos, C., Dimitriou, E. & Rogdakis, Y. (2003). Spatial patterns and temporal trends in the fishery landings of the Messolonghi-Etoliko lagoon system (western Greek coast). *Scientia Marina*,   67, 501-511.

Khongchai, N., Vibunpant, S., Eiamsa-ard, M. & Supongpan, M. (2003). Preliminary analysis of demersal fish assemblages in coastal waters of the Gulf of Thailand, In Silvestre G, Garces L, Stobutzki I, Ahmed M, Valmonte-Santos RA, Luna C, Lachica-Alino L, Munro P, et al (eds) Assessment, Management and Future Directions for Coastal Fisheries in Asian Countries. Worldfish Center Conference Proceedings 67, pp 249-262.

Legendre, P. (2005). Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 226-245.

Legendre, P., Borcard, D. & Peres-neto, P. (2005). Analyzing beta diversity: partitioning the spatial. *Ecological Monographs*, 75, 435-450.

Liengpornpan, S., Jaroensutasinee, M. & Jaroensutasinee, K. (2004). Habitat Characteristics of Croaking Gouramis (*Trichopsis Vittata*). Organizing Committee of 30th Congress on Science and Technology of Thailand. [Online] Available: http://www.scisoc.or.th/stt/30/sec_b/paper/stt30_B0040.pdf (May 10, 2008).

McDowall, R. M. (1995). Seasonal pulses in migrations of New Zealand diadromous fish and the potential impacts of river mouth closure. *New Zealand Journal of Marine and Freshwater Research*, 29, 517-526.

McNeil, D. R. & Tukey, T. W. (1973). Higher-order diagnosis of two-way tables, illustrated on two sets of demographic empirical distributions. *Biometrics*, 31, 487-510.

Musick, J. A., Harbin, M. M., Berkeley, S.A., Burgess, G. H., Eklund, A. M., Findley, L., Gilmore, R. G., Golden, J. T., et al. (2001). Marine, Estuarine, and Diadromous Fish Stocks at Risk of Extinction in North   America (Exclusive of Pacific Salmonids). *Fisheries*, 25, 6-30.

Ahmadi-Nedushan, B., St-Hilaire, A., Bérubé, M., Robichaud, E., Thiémonge, N. & Bobée, B. (2006). A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Research and Applications*, 22, 503-523.

National Institute of Coastal Aquaculture (NICA). (2007). Annual Report 2006. National Institute of Coastal Aquaculture. Department of Fisheries, Ministry of Agriculture and Co-operatives, Thailand. pp 1-102.

Ratanachai, C. & Kanchanasuwan, W. (2005). Master Plan for Songkhla Lake Basin Management, Executive Summary (Team leader). Final report, Prince of Songkla University, Songkhla, Thailand, pp 1-59.

Salonen, E. & Mutenia, A. (2007). Alien fish species in northernmost Finland. Helsinki. Finfish Game and Fisheries Research Institute, pp 1-16.

Theil, H. (1983). *Linear Algebra and Matrix Methods in Econometrics*. North-Holland Publishing Company, pp 1-64.

Thompson, R. M. & Townsend, C. R. (1999). The effect of seasonal variation on the community structure and food-web attributes of two streams: implication for food-web science. *OIKOS,* 87, 75-88.

Venables, W. N. & Smith, D. M. The R Development Core Team. (2004). An Introduction to R: Notes on R: A Programming Environment for Data Analysis and Graphics Version 2.6.2 (2008-2-08). [Online] Available: http://cran.r-project.org/doc/manuals/R-intro.pdf. (March 3, 2008)

Vaughan, I. P. & Ormerod, S. J. (2005). Increasing the value of principal components analysis for simplifying ecological data: a case study with rivers and river birds. *Journal of Applied Ecology*, 42: 487-497.

Welcomme, R. L., Winemillerb, K. O. & Cowx, I. G. (2006). Fish environmental guilds as a tool for assessment of ecological condition of rivers. *River Research and Applications*, 22, 377-396.

Wetzel, R.G. (2001). *Limnology. Lake and River Ecosystems*. (3rd ed.). San Diego: Academic Press.

Winemiller, K. O. & Jepsen, D. B. (1998). Effects of seasonality and fish movement on tropical river food webs. *Journal of Fish Biology*, 53, 267-296.

Wirth, F. F., Dugger, D. M. & Creswell, L. (2004). Commercial Scale Penaeid Shrimp Demonstration in Inland Freshwater Systems. Final Project Report for Cost Reimbursable Contract 007199 between FL DACS, Division of Aquaculture and University of Florida, IFAS. [Online] Available: http://Floridaaquaculture.com/publications/Penaeid%20Shrimp%20Report.pdf (February 25, 2008)

Yanez-Arancibia, A., Dominguez, A. L. L. & Pauly, D. (1994). Coastal Lagoons as Fish Habitats, In Kjerjve B (ed) Coastal lagoon Processes. Amsterdam: Elservier Science Publishers, pp 363-376.

Table 1. Species disconnected from a single cluster

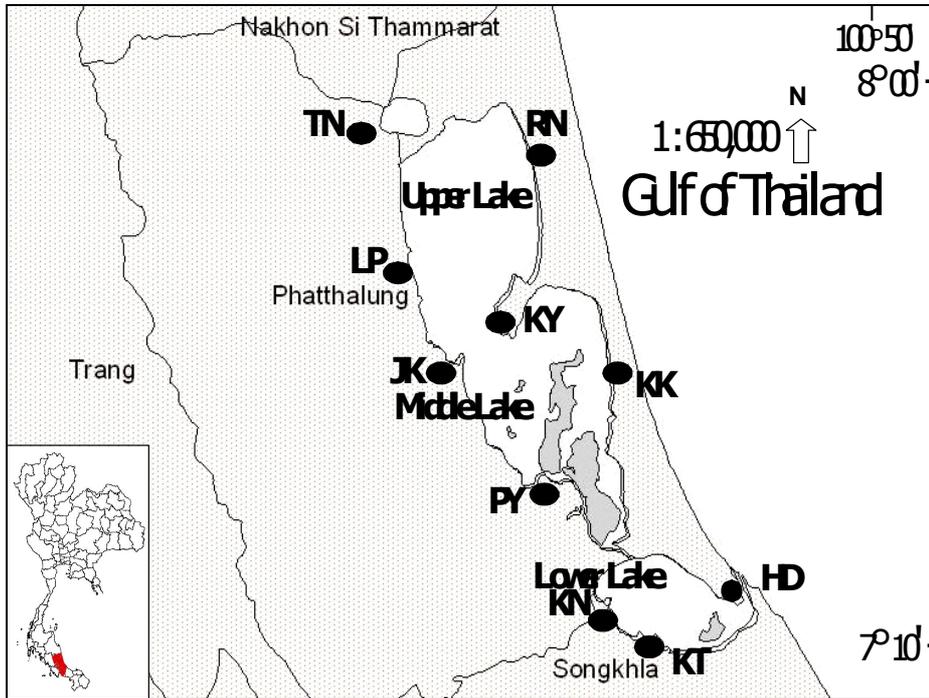| Label	Scientific name	Common name	family	Category
A	*Penaeus monodon*	Black tiger shrimp	Penaeidae	Marine invertebrate
B	*Anodontostoma chacunda*	Chacunda gizzard shad	Clupeidae	Marine vertebrate
C	*Hypoatherina valenciennei*	Sumatran silverside	Atherinidae	Marine vertebrate
D	*Epinephelus coioides*	Greasy grouper	Serranidae	Marine vertebrate
E	*Stolephorus sp*	Broadhead anchovy	Engraulidae	Estuarine vertebrate
F	*Trichogaster pectoralis*	Snakeskin gourami	Osphronemidae	Freshwater vertebrate
G	*Portunus pelagicus*	Blue swimming crab	Potunidae	Marine invertebrate
H	*Cyclocheilichthys apogon*	Beardless barb	Cyprinidae	Freshwater vertebrate
I	*Eleutheronema tetradactylum*	Fourfinger threadfin	Polynemidae	Marine vertebrate
J	*Puntius partipentazona*	Sumatran tiger barb	Cyprinidae	Freshwater vertebrate
K	*Escualosa thoracata*	Thai anchovy	Clupeidae	Estuarine invertebrate
L	*Johnius dussumieri*	Sin croaker	Sciaenidae	Marine vertebrate
M	*Abalistes stellaris*	Starry triggerfish	Balistidae	Marine vertebrate
N	*Mystus cavasius*	Black lancer catfish	Bagridae	Freshwater vertebrate
O	*Lactarius lactarius*	False trevelly	Lactariidae	Marine vertebrate
P	*Pseudorhombus arsius*	Large tooth flounder	Paralichthyidae	Marine vertebrate
Q	*Lutjanus russellii*	Russell's snapper	Lutjanidae	Marine vertebrate
R	*Oreochromis niloticus X O. Mossambicus*	Red tilapia (hybrid tilapia)	Cichlidae	Marine vertebrate
S	*Acantopsis choirorhynchos*	Horse face roach	Cobitidae	marine vertebrate
T	*Scomberomorus commerson*	Spanish mackerel	Scombridae	Marine vertebrate
U	*Caranx sexfasciatus*	Dusky Jack	Carangidae	Marine vertebrate
V	*Metapenaeus ensis*	Greasy back shrimp	Penaeidae	Estuarine invertebrate
W	*Parapenaeopsis hardwickii*	Spear shrimp	Penaeidae	Marine invertebrate
X	*Glossogobius aureus*	Sand goby, Golden goby	Gobiidae	Estuarine vertebrate
Y	*Oreochromis mossambicus*	Mozambique tilapia	Cichlidae	Marine vertebrate
Z	*Trichopsis vittata*	Croaking gourami	Osphronemidae	Freshwater vertebrate
a	*Tetraodon nigroviridis*	Spotted green pufferfish	Tetraodontidae	Marine vertebrate
b	*Sepioteuthis lessoniana*	Cuttlefish	Loliginidae	Marine invertebrate
c	*Aseraggodes dubius*	Largescale tonguesole	Soleidae	Marine vertebrate
d	*Bregmaceros mcclellandi*	Spotted codlet	Bregmacerotidae	Marine vertebrate
e	*Scomberomorus guttatus*	Indo-pacific mackerel	Scombridae	Marine vertebrate
f	*Pomadasys kaakan*	Lined silver grunt	Haemulidae	Marine vertebrate
g	*Macrobrachium rosenbergii*	Giant freshwater prawn	Palaemonidae	Freshwater invertebrate
h	*Macrobrachium equidens*	Dwarf prawn	Palaemonidae	Estuarine invertebrate
i	*Metapenaeus ensis*	Bird shrimp	Penaeidae	Estuarine invertebrate
j	*Acetes sp*	Acetes	Sergestidae	Estuarine invertebrate
k	*Cloridopsis dubia*	Mantis shrimp	Squillidae	Marine invertebrate
l	*Penaeus merguiensis*	Banana prawn	Penaeidae	Marine invertebrate
m	*Penaeus semisulcatus*	Green tiger prawn	Penaeidae	Marine invertebrate
n	*Metapenaeus lysianassa*	Small white shrimp	Penaeidae	Marine invertebrate
o	*Parapocryptes serperaster*	Pink borrowing goby	Gobiidae	Marine vertebrate
p	*Triacanthus biaculeatus*	Short-nosed tripodfish	Triacanthidae	Marine vertebrate
q	*Oxyurichthys microlepis*	Smallscaled goby	Gobiidae	Estuarine vertebrate
r	*Terapon jarbua*	Thornfish	Terapontidae	Marine vertebrate
s	*Trichiurus lepturus*	Largehead hairtail	Trichiuridae	Marine vertebrate
t	*Uroteuthis duvauceli*	Indian squid	Loliginidae	Marine invertebrate
u	*Metapenaeopsis stridulans*	Fiddler shrimp	Penaeidae	Marine invertebrate
v	*Metapenaeus brevicornis*	Yellow shrimp	Penaeidae	Marine invertebrate
w	*Octopus dollfusi*	Dollfus' Octopus	Octopodidae	Marine invertebrate
x	*Leiognathus brevirostris*	Short-nosed ponyfish	Leiognathidae	Estuarine vertebrate

Figure 1. Map of Songkhla Lake and ten fish catch landing sites for data collection
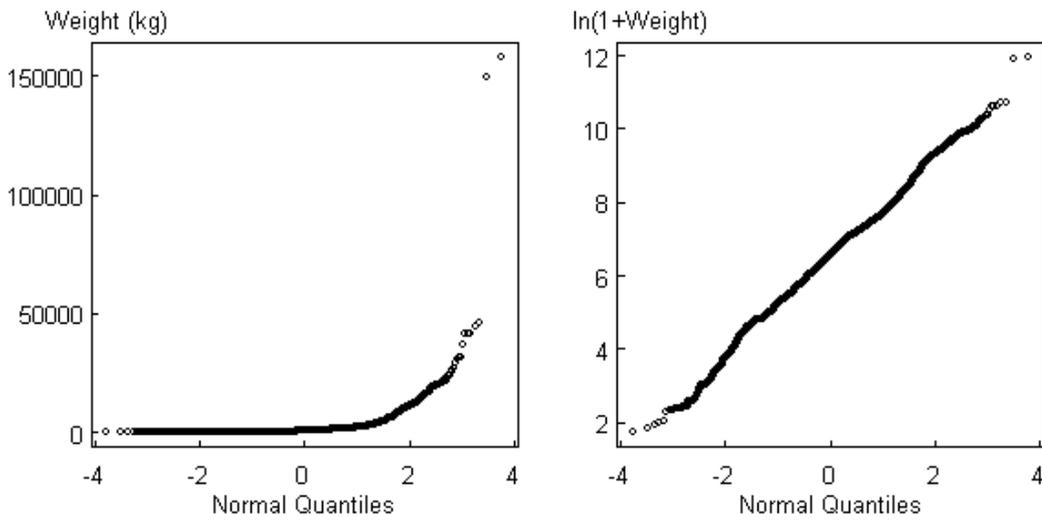

Figure 2. Distributions of monthly catch weights before and after data transformation
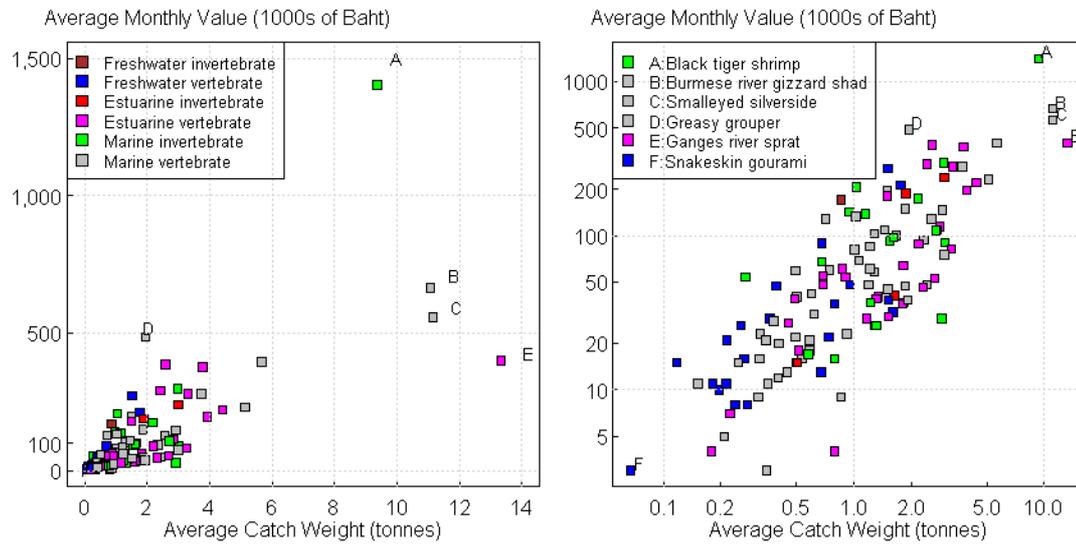
Figure 3. Scatter plots of average monthly catch weights and monetary value for each species
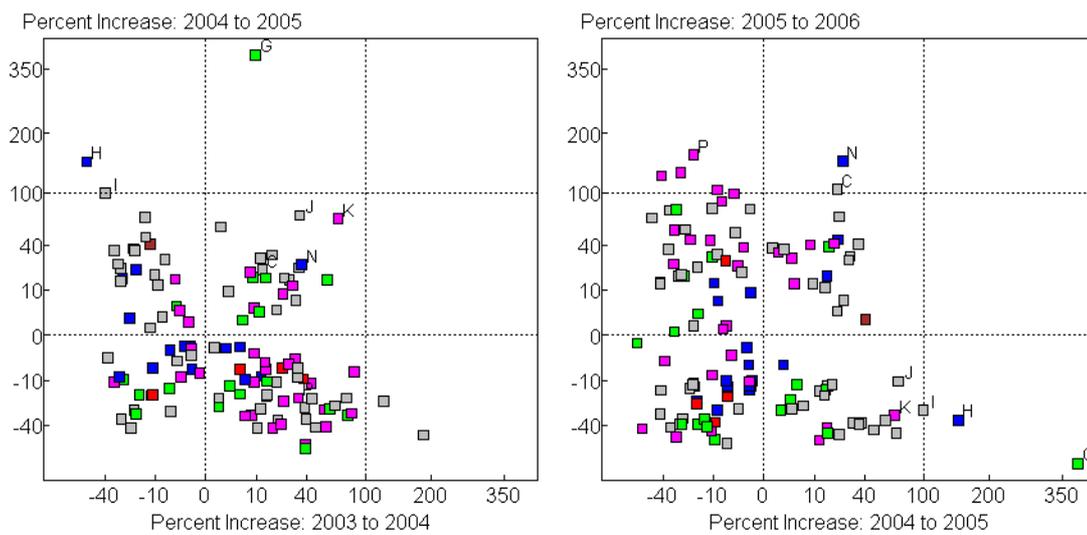

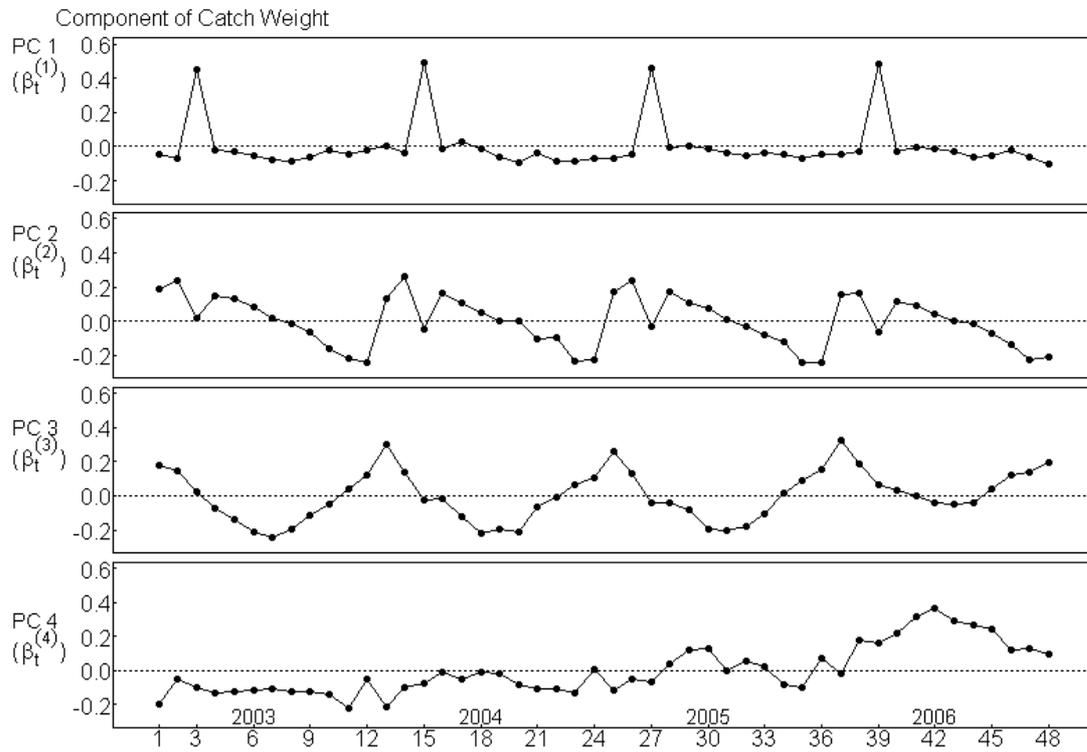Figure 4. Annual catch weight changes for each species from 2003-2004, 2004-2005 and 2005-06

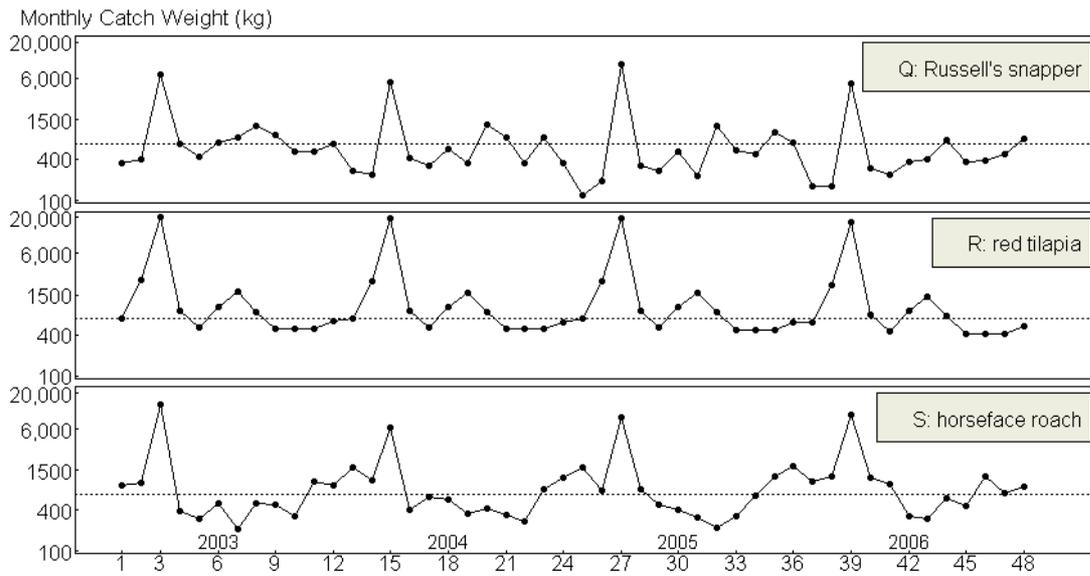Figure 5. Plot of first four principal components of catch weights

Figure 6. Monthly catch weights of species characterized by first model component

Figure 7. Monthly catch weights of species containing second model component
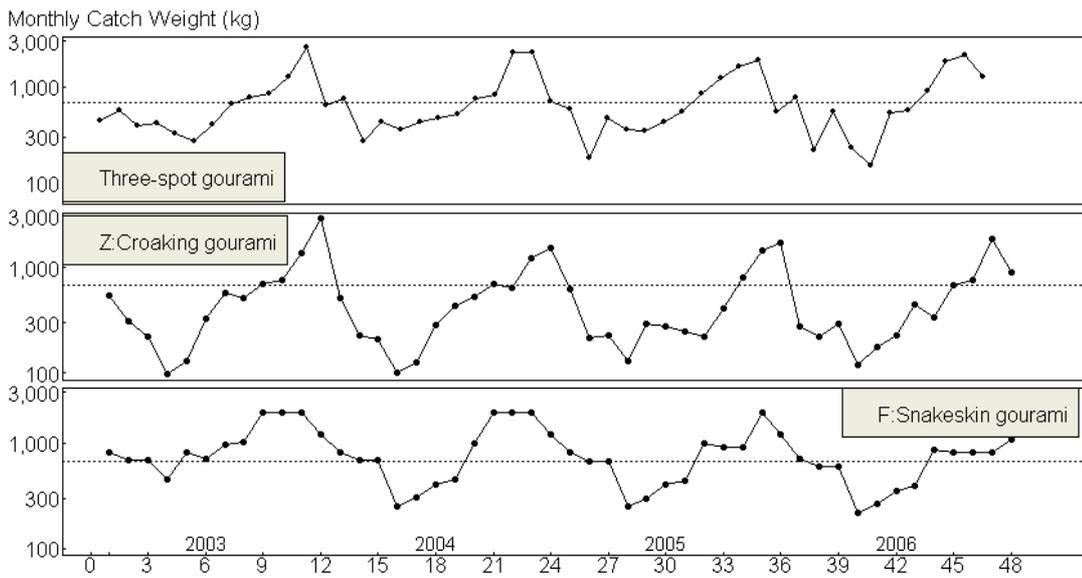

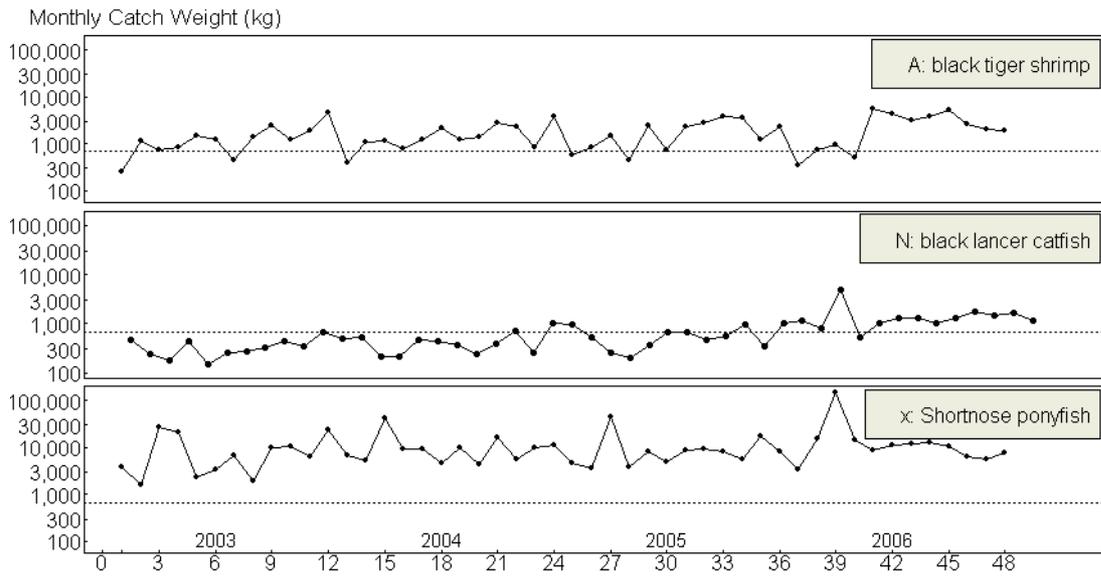Figure 8. Monthly catch weights of species containing third model component

Figure 9. Monthly catch weights of species containing forth model component
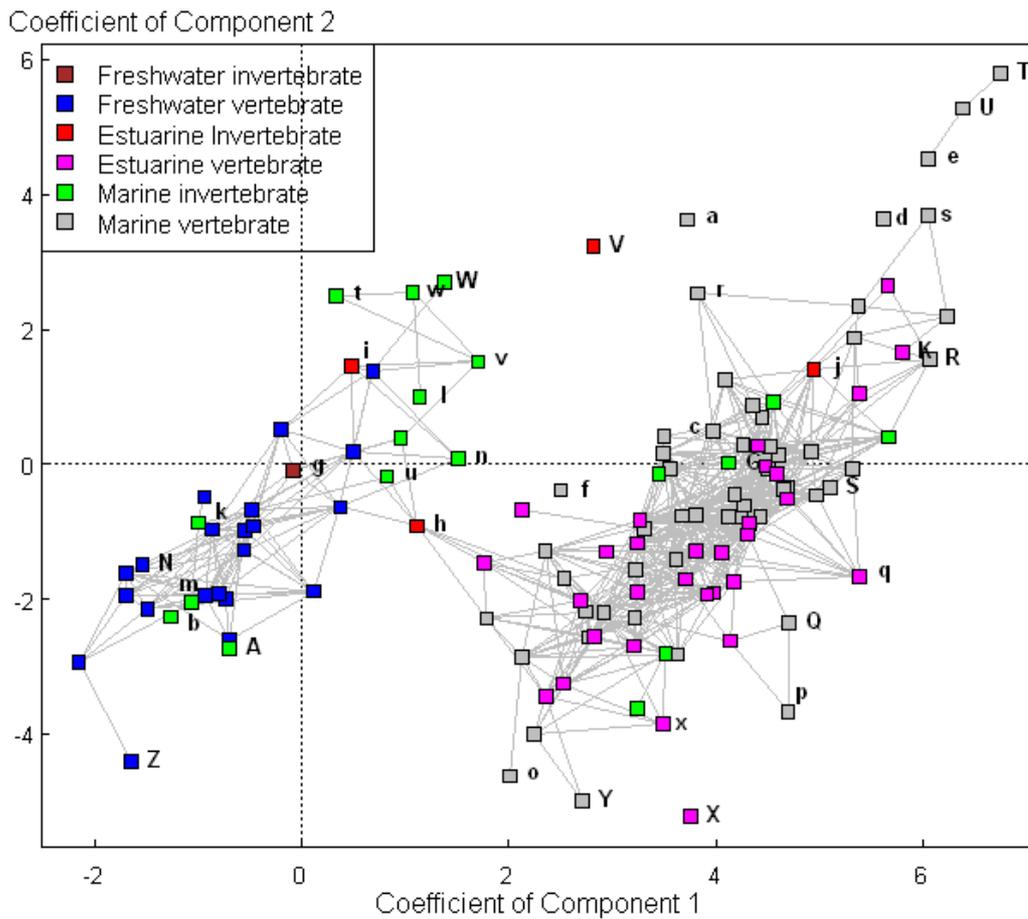

Figure 10. Scatter plots of regression coefficients of principal components