Chapter 4

Models for clustering fish community structure by species

In this chapter we present some preliminary analysis and further analysis using statistical regression models to the time series of monthly fish catch weights over the period from January 2003 to December 2006. The analysis is based on well known statistical methods, although some of these methods are new to biological science. The results presented in this chapter also appear in two papers, the first of which is published (Chesoh et al 2009) and is reproduced in Appendix III, and the second of which has been submitted for publication.

4.1 Characteristics of the data

Species composition and diversity

Monthly fish catch weights in Songkhla Lake over the period January 2003 to December 2006 shows a total of 68 families belonging to 127 species (see Appendix I). The ten families with the highest percent of catch weights, constituting 62.3% of the total annual landing, are Penaeidae (24.4%), Leiognathidae (18.9%), Gobiidae (10.6%), Clupeidae (9.6%), Ariidae (9.3%), Cyprinidae (6.7%), Mugillidae (5.9%), Bragridae (5.0%), Paleamonidae (4.9%), and Chandidae (4.8%).

The 127 species are defined in terms of their biological and habitat characteristics. There are 53 marine vertebrates; 30 estuarine vertebrates; 21 freshwater vertebrates; 4 estuarine invertebrates; 18 marine invertebrates, and one freshwater invertebrate. The largest proportion of catches was estuarine vertebrates representing about 34.2% of the mean annual catch weight, followed by marine vertebrates with 27.0%, marine invertebrates with 15.3%, and estuarine invertebrates with 11.2%. In all these first four fish groups the highest catches occurred between March and May. In contrast, freshwater vertebrates, representing about 10.9% of the mean annual catch weight, and freshwater invertebrates with 1.4%, showed a steady catching rate but the mean annual catch of the freshwater invertebrate appeared to slightly decrease in the most recent two years.

Annual changes in catch weights from 2003-2004, 2004-2005 and 2005-2006 for each species are shown in Figure 4.1.

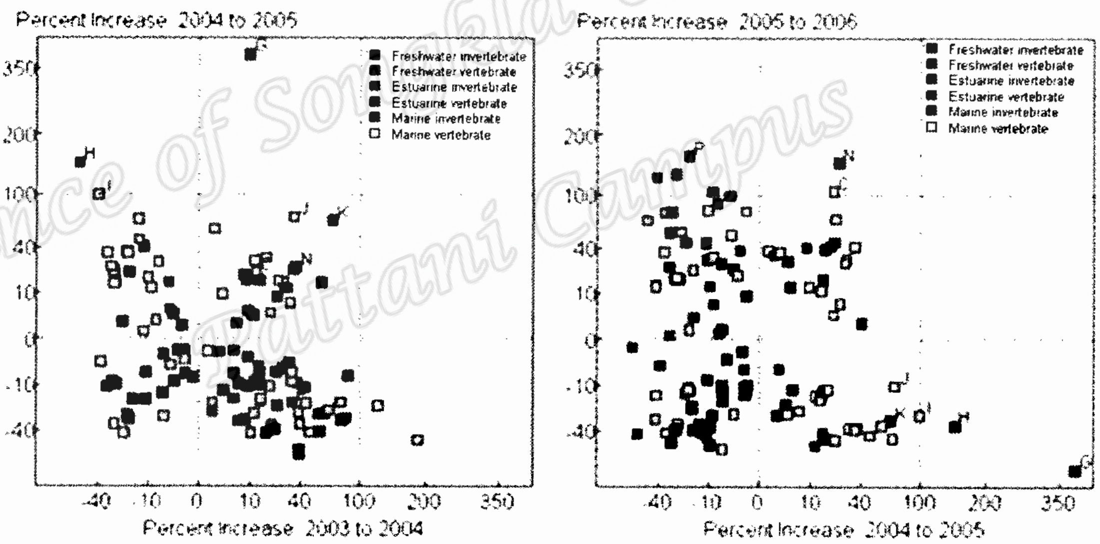

Figure 4.1 Annual catch weight changes for each species from 2003-2004, 2004-2005 and 2005-2006

In Figure 4.1 shows that most estuarine (green) and marine (magenta) invertebrates exhibited an overall decline in catch weight, as did freshwater fish (blue and one brown colour), while estuarine (red) and marine (grey) vertebrates showed a steady change over the 4 year period.

Catch by fishing gear

There are three major types of fishing gear that are commonly used in Songkhla Lake, namely set-bag net, trap and gill nets. The proportion of catch weight classified by these three types of fishing gear is shown in Figure 4.2.

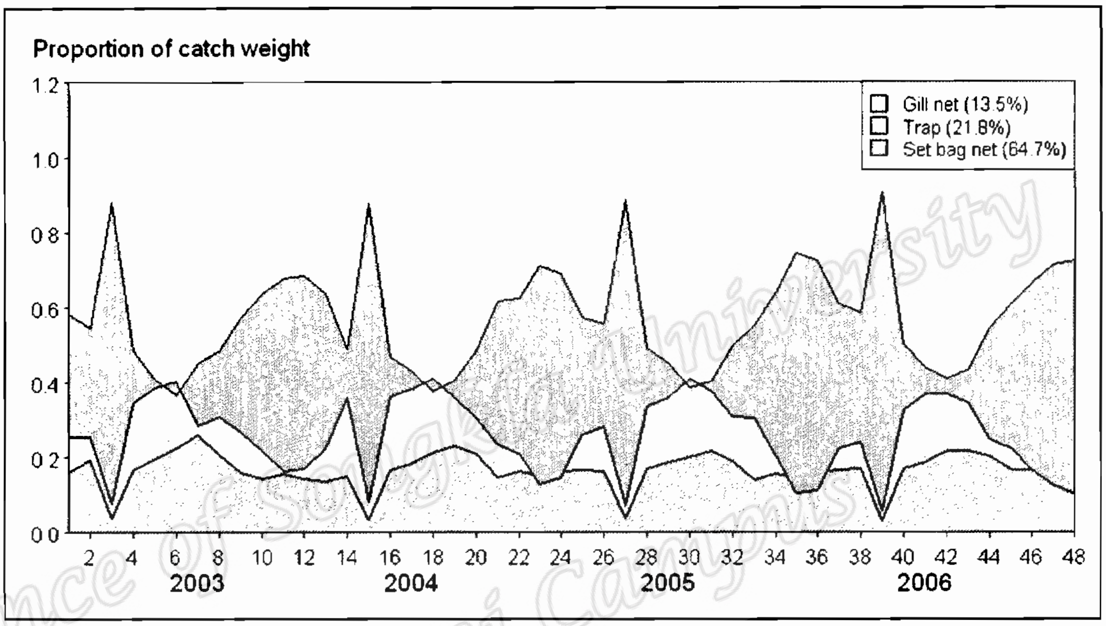

Figure 4.2: Proportion of catch weights categorized by three fishing gears

Over the 4 year period, 64.7% of the fish were caught using set-bag nets, 21.8% using traps and 13.5% using gill nets. The set bag nets were used mostly during March, while the traps were used more frequently from April to June.

The average percentage of the six biological/habitat groups classified by three types of gears during 2003-2006 is shown in Table 4.1.

Table 4.1 Total catch weights (tonnes) and average percentage of 6 fish groups caught by 3 types of gear during 2003-2006

Gear year	Freshwater invertebrate (1 sp*.)	Freshwater vertebrate (21 sp.)	Estuarine invertebrate (4 sp.)	Estuarine vertebrate (30 sp.)	Marine invertebrate (18 sp.)	Marine vertebrate (53 sp.)	Total (127 sp.)
Set bag net							
2003	8.8	85.3	134.1	78.1	639.2	572.4	1,517.9
2004	6.9	74.4	147.7	74.6	789.9	608.8	1,702.4
2005	8.6	64.2	190.5	72.8	681.5	582.7	1,600.4
2006	6.9	54.0	108.6	70.0	839.8	572.3	1,651.5
mean	*7.8*	*69.5*	*145.2*	*73.9*	*737.6*	*584.1*	*1,618.1*
Trap							
2003	21.8	179.4	211.3	39.7	43.5	36.4	532.2
2004	19.8	225.3	222.5	38.5	50.2	38.2	594.5
2005	29	200.5	218.4	38.3	48.8	39.9	574.9
2006	30.7	137.4	174.6	37.7	53.7	39.2	473.4
mean	*25.3*	*185.7*	*206.7*	*38.6*	*49.1*	*38.4*	*543.8*
Gill net							
2003	2.3	25.4	26.8	178.1	57.7	47.8	338.1
2004	2.1	25.8	29.6	172.5	66.7	49.4	346.1
2005	2.9	22.0	35.9	170.1	61.9	51.0	343.7
2006	3.4	16.2	22.9	165.4	66.7	49.8	324.4
mean	*2.7*	*22.4*	*28.8*	*171.5*	*63.3*	*49.5*	*338.1*
Total mean	*35.8*	*277.6*	*380.7*	*284.0*	*850.0*	*672.0*	*2500.0*
%	1.4	11.1	15.2	11.4	34.0	26.9	100.0

The mean annual fish catch weight in Songkhla Lake was 2,500 tonnes with a range of 2,388.2–2643.0 tonnes. The highest catch weight was seen in 2004 and the lowest was seen in 2003 for all three types of gear.

4.2 Statistical modelling using species and months as predictors

We modeled the natural logarithm of catch weights of 127 fish species in 48-monthly collections from 2003 to 2006. The data thus comprise 6,096 records (127 species × 48 months). Using equation (2.7) (see details in Chapter 2), the eigenvectors for catch weight corresponding the largest four eigenvalues of the data covariance matrix (adjusted for species) were calculated. The four components $\{\beta_t^{(k)}\}$, k =1, 2, 3 and 4, fit the data reasonably well with r-squared equal to 0.877. The distribution of eigenvector plots of each of four components $\beta_t^{(k)}$, k = 1 to 4, is shown in Figure 4.3.

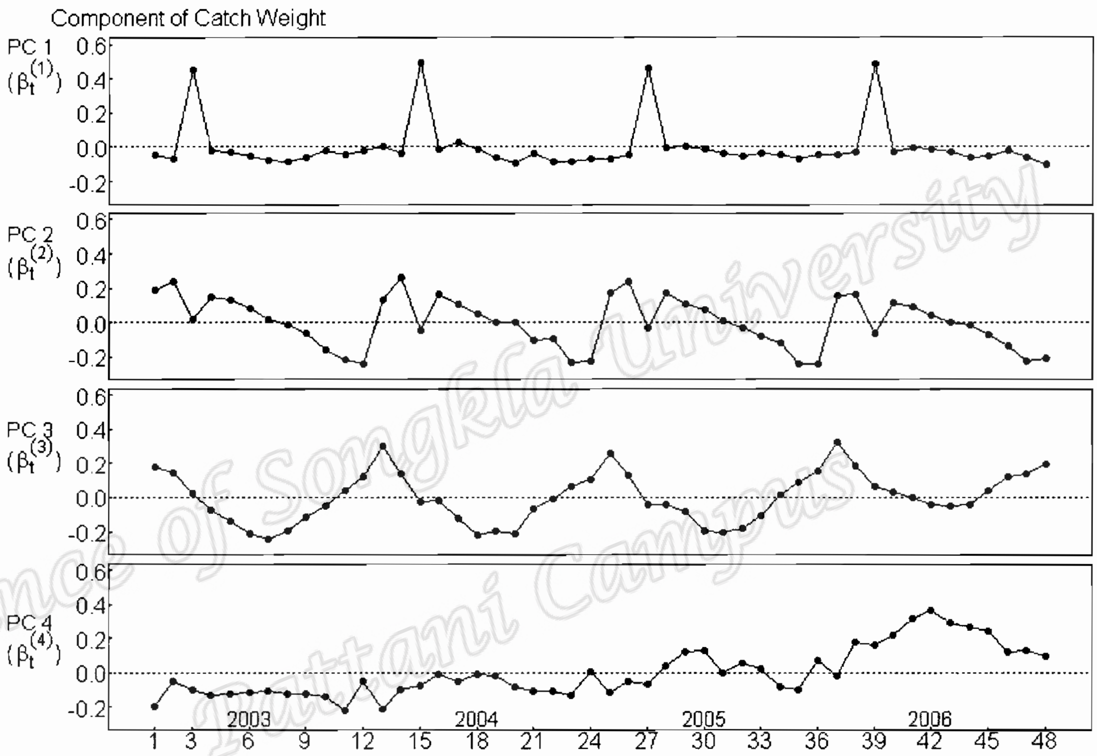

Figure 4.3 Plot of first four principal components of catch weights

The first eigenvector component shows a similar seasonal pattern for the whole period, with a spike occurring in March of each year. The second component shows a regular seasonal pattern with peaks occurring in February and declining to a minimum in December of each year after dipping in March. The third component also shows a seasonal pattern for the whole period with peaks occurring in January. The fourth component is less regular and shows an upward trend that rises more sharply in the most recent year of the study period. The monthly catch weights of some example species characterized by these four components are shown in Figures 4.4–4.7.

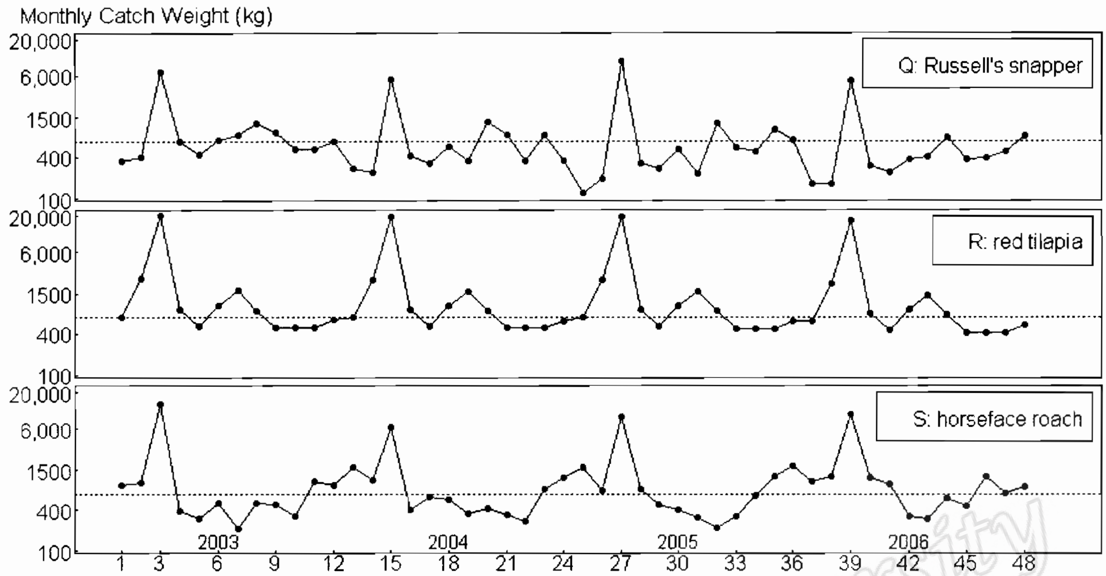

Figure 4.4 Monthly catch weights of species characterized by first model component

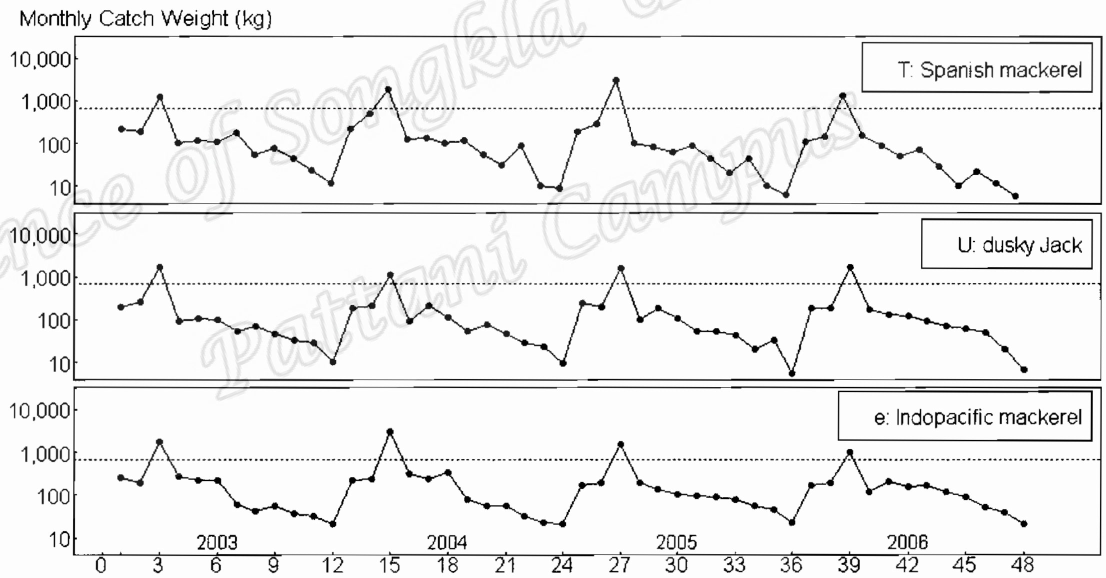

Figure 4.5 Monthly catch weights of species containing second model component

We see that the first component is characterized by species such as Russell's snapper (*Lutjanus russellii*), Red tilapia and Horseface roach (*Acantopsis choirorhynchos*). Figure 4.5 shows that 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*).

However, these three species have maximum catch weights in March, indicating sufficient presence of the first component to offset the dip in March of the second component.

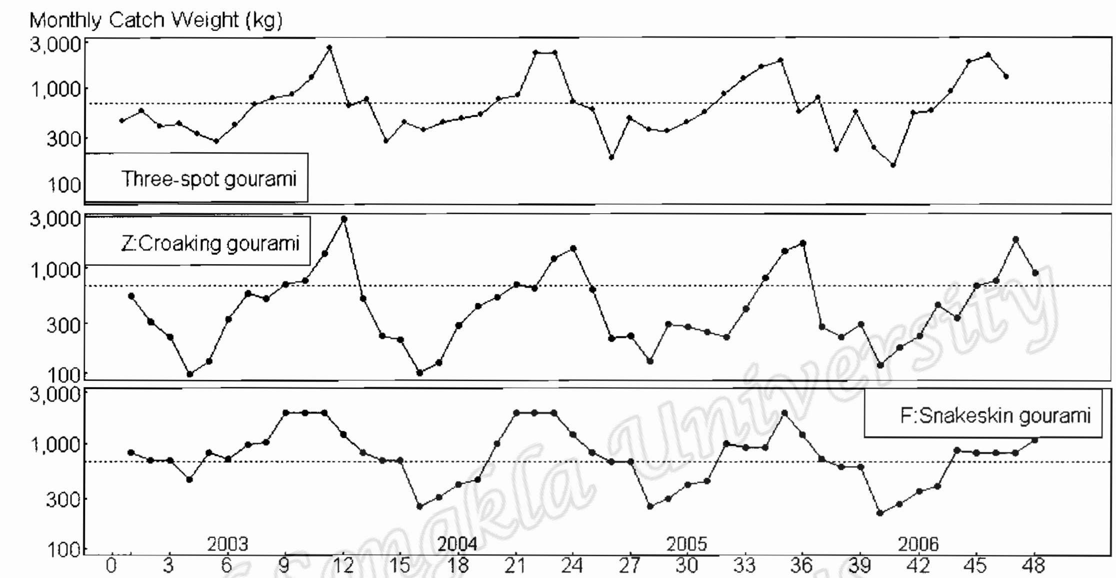

Figure 4.6 Monthly catch weights of species containing third model component

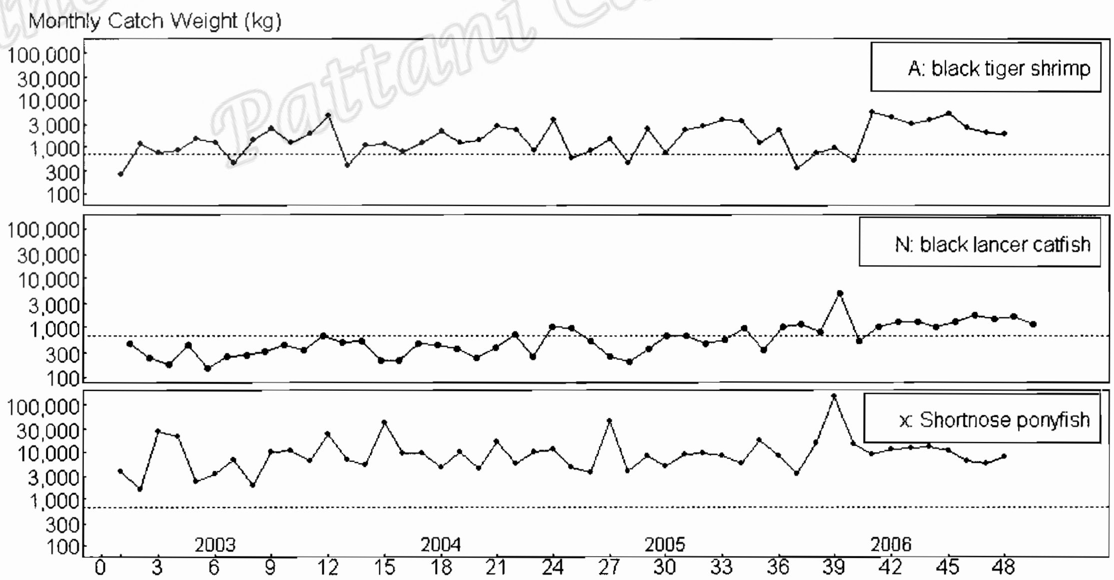

Figure 4.7 Monthly catch weights of species containing forth model component

In Figure 4.6, the three species most characterized by the third component pattern are the 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 4.7 shows that three species that follow the pattern of the fourth component are the black tiger prawn (*Penaeus monodon*), black lancer catfish (*Mystus cavasius*), and shortnose ponyfish (*Leiognathus brevirostris*), with a slightly increasing trend in 2006.

In contrast to the regular seasonal cycles exhibited by the first two components, the third and fourth components are less regular and show an upward trend that rises more sharply in 2006. A scatter plot of the first two component coefficients is shown in Figure 4.8. In this graph, 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). Two distinct clusters are clearly visible. The smaller cluster on the left hand side of Figure 4.8 contains all freshwater species of vertebrates (blue squares) and the single freshwater invertebrate (the giant freshwater prawn, labeled g), whereas the second group contains all marine vertebrates (grey squares) and invertebrates (green). The estuarine species appear in both clusters.

Figure 4.8 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 Appendix III, and includes all species not connected to other species of the same type within the same cluster.

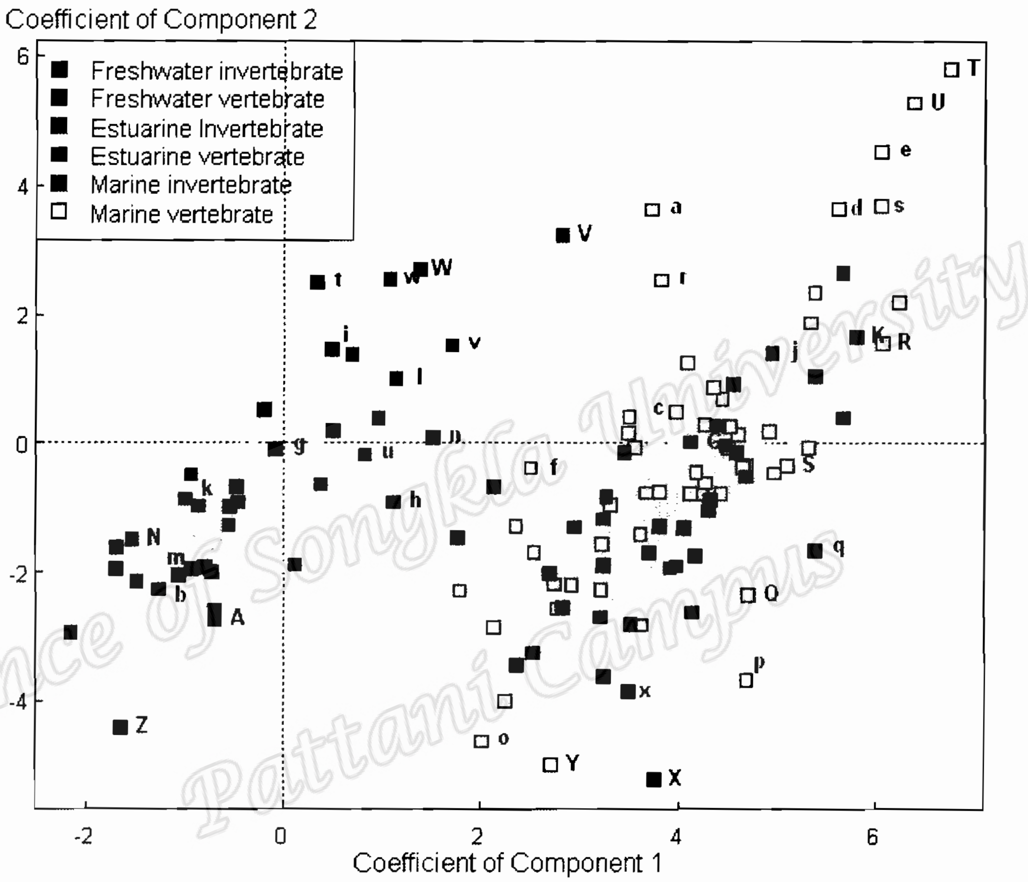

Figure 4.8 Scatter plot of first two components in the statistical model

4.3 Statistical modeling using species and season-gear as predictors

In this model the same data are used but gear is also taken into account. To substantially reduce the large number of parameters in this model, the period factor comprising 48 months was replaced simply by 6 seasonal (bimonthly) parameters, with year removed from the model, so that annual trends may be seen in residuals rather than in the model parameters. The swamp eel was excluded from this analysis because it cannot be caught by a gill net. Thus the data comprise 18,144 records (126 species × 4 years × 12 months × 3 gear types). The seasonal fishing period variable is thus defined as the combination of two adjacent months, January and February, March and April, and so on, giving a total of six seasonal fishing periods. To reduce the number of factors in the model, a season-gear variable was defined as the combination of seasonal fishing period and gear (6 bi-monthly periods × 3 gear types). Thus the season-gear comparisons of fish abundance in terms of catch weight data reduce to 2,268 records (126 species × 18 season-gear parameters). Using equation (2.7), the eigenvectors for catch weight were calculated according to species and season-gear. The model shows that three components $\{\beta_t^{(k)}\}$, k =1, 2, and 3, fit the data well with r-squared equal to 0.923, although the residuals plot indicates some departure from the statistical normality assumption as shown in Figure 4.9.

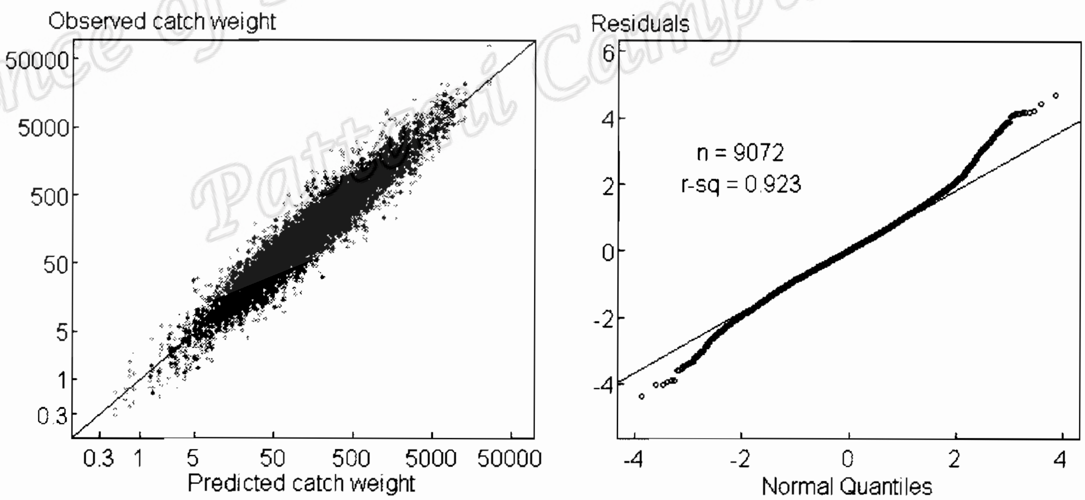

Figure 4.9 Plots of observed against predicted catch weight (left panel), and residuals against corresponding normal quantiles (right panel).

The coefficients and standard errors from the multiplicative linear model are shown in Appendix II. The first three seasonal catch components ($\beta_t^{(1)}$, $\beta_t^{(2)}$, $\beta_t^{(3)}$) are shown in Figure 4.10.

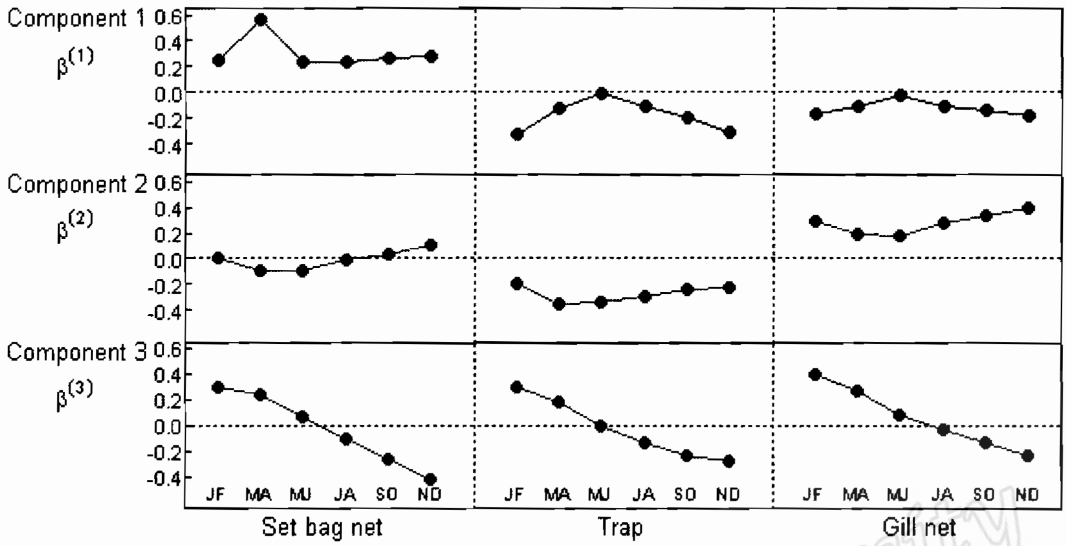

Figure 4.10 First three season-gear components for fish catch during 2003-2006.

In Figure 4.10, the first component shows a high catch pattern for the whole period, with a set bag net spike occurring in March-April of each year, and trap and gill net spikes occurring in May-June of each year. The first component also reflects the fact that using set bag nets resulted in the highest fish catches, while the lowest catches occurred when traps and gill nets were used.

The second component shows an interaction of regular seasonal and gear catch pattern with a slight decline occurring from January-February to May-June of each year, followed by a gradual increase until November-December, for each gear type. The curves indicate that highest catch weights occurred in gill nets while the lowest occurred in traps.

The third component shows that for all three gear types there was a similar strong seasonal pattern with steady decline occurring from January-February and further decline to a low in November-December of each year. Also the patterns for the three types of gear were very similar.

We can use the coefficients from the model (Appendix II) to get the model for any specified species. For example, Brownback trevally (species code 114) has a mean (natural logarithmic) catch weight of 4.05, and the set of coefficients $\left\{\alpha_s^{(1)}, \alpha_s^{(2)}, \alpha_s^{(2)}\right\}$ equal to {5.35, 0.04, 0.04}, so the model is given by

$$y_{st} = \mu_s + \alpha_s^{(1)}\beta_t^{(1)} + \alpha_s^{(2)}\beta_t^{(2)} + \alpha_s^{(3)}\beta_t^{(3)},$$

that is,

$$y_{114,t} = 4.05 + 5.35\beta_t^1 + 0.04\beta_t^2 + 0.04\beta_t^3$$

Then we plot the bi-monthly catch weights for each gear type and species component pattern together with the model (red line), as shown in Figure 4.11. Examples of three more species, Largescale archerfish, Stork shrimp and Dusky jack, are also shown in this figure.

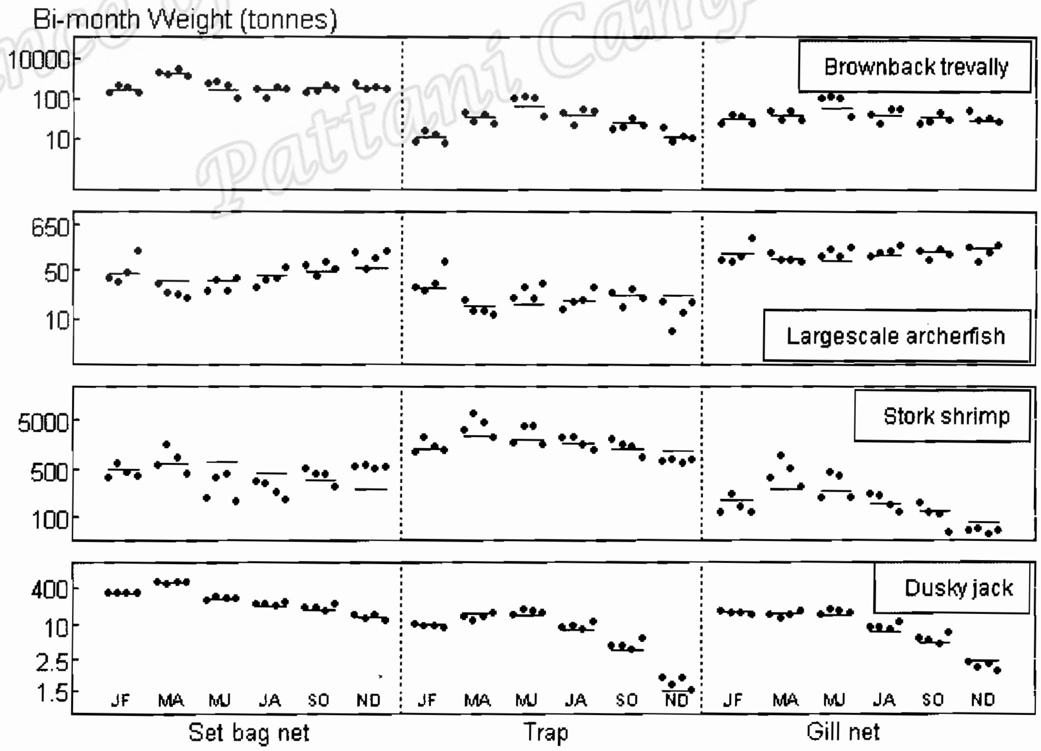

Figure 4.11 Bi-monthly catch weights by gear type showing four typical species following the dominant patterns from 2003 to 2006

Figure 4.11 shows that the Brownback trevally, Largescale archerfish, Stork shrimp, and Dusky jack represent the first, second and third components, respectively.

Plots of the three components showing the clustering of fish species are shown in Figure 4.12.

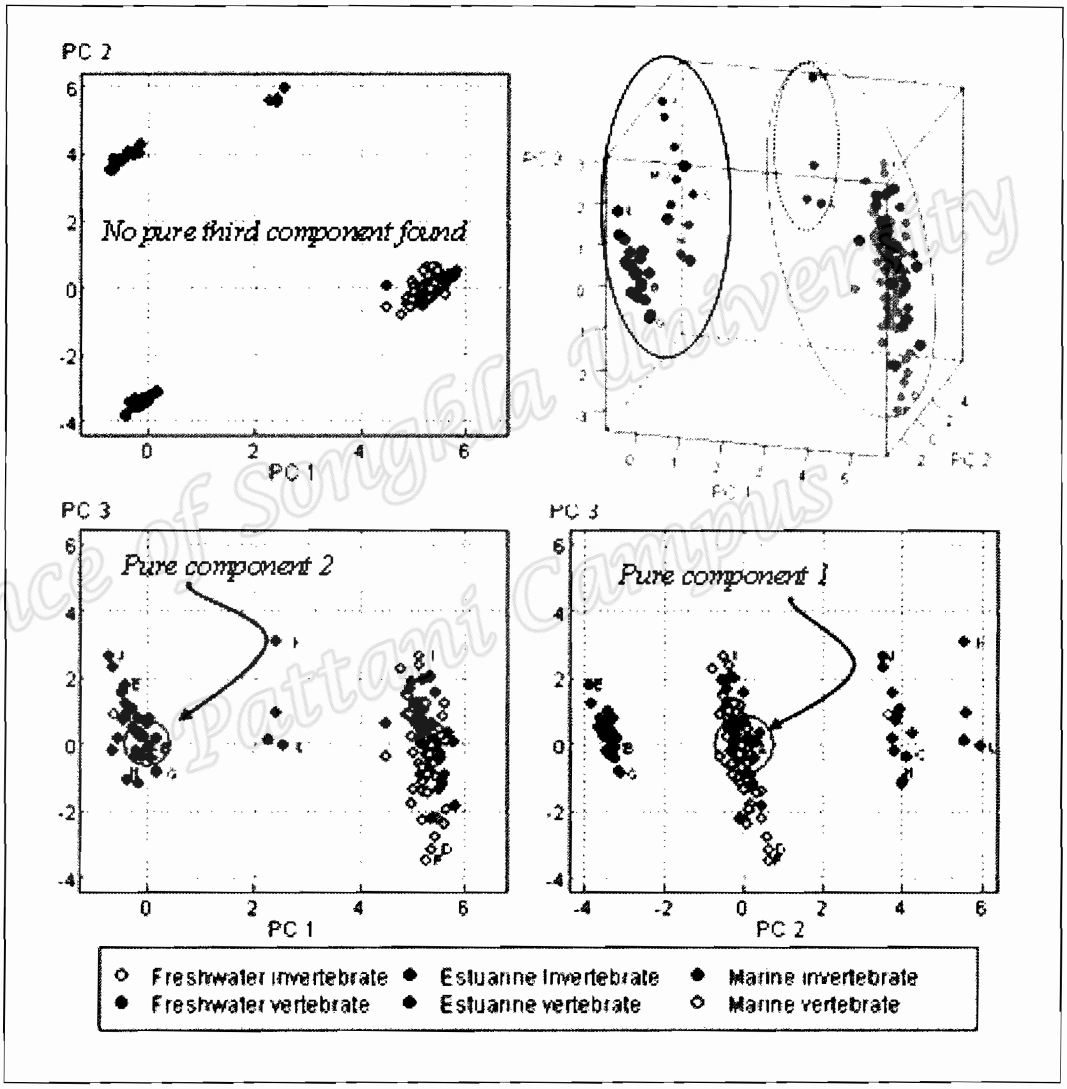

Figure 4.12 Three principal components showing clustering of fish species

Figure 4.12 shows two distinct fish community clusters. The dominant cluster contains all 83 marine and estuarine vertebrate species, ten species of marine invertebrates and one of estuarine invertebrate (bird shrimp, *Metapenaeus ensis*). Two distinct sub-clusters were also seen. The first sub-cluster represents most species of estuarine and marine vertebrates, and the second is a small group of three offshore fish species, namely Indopacific mackerel (*Scomberomorus guttatus*), streaked seerfish (*Scomberomorus lineolatus*) and bigeye trevally (*Caranx sexfasciatus*), and one isolated species of cuttlefish (*Sepioteuthis lessoniana*). The less-dominant cluster includes twenty-one of all species of freshwater vertebrates and eleven species of estuarine and marine invertebrates.

4.4 Summary

The catch weights of 127 fish species in 48-monthly collections from 2003 to 2006 in Songkhla Lake was found to have a mean of 2,499.9 tonnes (range 2,388.2 – 2,643.0). These fish were caught by three major types of fishing gear, namely set bag net (64.7% of catch weight), followed by traps (21.8%) and gill nets (13.5%). The data were categorized in terms of biological and habitat characteristics into 6 groups, marine vertebrate, estuarine vertebrate, freshwater vertebrate, estuarine invertebrate, marine invertebrate, and freshwater invertebrate. During the 4 year period, the catch weights of freshwater and estuarine fish increased, while that of marine invertebrates decreased. Although several species are ubiquitous in all habitats, some are anadromous or catadromous species and so have specific habitats in Songkhla Lake.

We fitted a linear regression model based on the natural logarithm of catch weights classified by species and month, using these factors as multiplicative determinants based on principal components, thus enabling assessment of species clustering. The model was found to have four components corresponding to predominant seasonal time series patterns. The r-squared value was 87.7%. Purely seasonal patterns were identified for the first two components, namely estuarine and marine vertebrates, and showed considerable seasonal fluctuations but otherwise appeared to be steady over the four year period.

A second linear regression model was fitted based on the natural logarithm of catch weights classified by species and season of the year after taking gear into account. Catch weights were first aggregated by species and combination of bi-monthly season and catching gear. The regression model had three component species and season-gear interaction terms and the r-squared was 92.3%. The first component contained most species of estuarine and marine vertebrates as well as some invertebrates, reflecting the fact that set bag net was the gear that resulted in the highest fish catch. The second component mainly contained freshwater fish and some marine invertebrates, and reflects the fact that highest catch for these species occurred in gill nets. The third component reflects on the seasonal fluctuation in catch weights.