www.trjfas.org
ISSN 1303-2712
Turkish Journal of Fisheries and Aquatic Sciences 11: 177-184 (2011)
DOI: 10.4194/trjfas.2011.0201

PROOF

A Model for Clustering Fish Community Structure with Application to Songkhla Lake Bi-monthly Catches 2003-2006

Sarawuth Chesoh[1], Chamnein Choonpradub[2,]*

[1] Ministry of Agriculture and Co-operatives, Department of Fisheries, Pattani, 94160, Thailand.
[2] Prince of Songkla University, Faculty of Science and Technology, Department of Mathematics and Computer Science, Pattani Campus, 94000, Thailand.

* Corresponding Author: Tel.: +66.819 633775; Fax: +66.819 633775;
E-mail: chesoh.s@hotmail.com

Received 12 September 2009
Accepted 14 November 2010

Abstract

Monthly catch weights in Songkhla Lake were collected over the period from January 2003 to December 2006, for a total of 126 species. Catch weights were first aggregated by species and combination of bi-monthly season of year and catching gear (set bag net, trap, or gill net) and were log-transformed to remove skewness. A regression model containing three species-season/gear components was then used to predict these outcomes. The first component was represented by the most species of estuarine and marine vertebrates as well as some invertebrates and reflected the fact that set bag net was the gear that resulted in the highest catches. The second component mainly represented freshwater fish and some marine invertebrates, and reflected the fact that most of these species were caught by gill nets. The third component focused on the seasonal fluctuations in catch weight. Such models can provide further tools in understanding of fish community structure clusterings in fishery landings.

Keywords: Fish assemblage distribution, Regression model, Lake fishery

A Model for Clustering Fish Community Structure with Application to Songkhla Lake Bi-monthly Catches 2003-2006

Özet

Monthly catch weights in Songkhla Lake were collected over the period from January 2003 to December 2006, for a total of 126 species. Catch weights were first aggregated by species and combination of bi-monthly season of year and catching gear (set bag net, trap, or gill net) and were log-transformed to remove skewness. A regression model containing three species-season/gear components was then used to predict these outcomes. The first component was represented by the most species of estuarine and marine vertebrates as well as some invertebrates and reflected the fact that set bag net was the gear that resulted in the highest catches. The second component mainly represented freshwater fish and some marine invertebrates, and reflected the fact that most of these species were caught by gill nets. The third component focused on the seasonal fluctuations in catch weight. Such models can provide further tools in understanding of fish community structure clusterings in fishery landings.

Anahtar Kelimeler: Fish assemblage distribution, Regression model, Lake fishery

Introduction

Fish assemblages and community structure are accepted as sensitive bioindicators of diversity index, habitat disturbances, environmental contamination, and ecological health (Dufrêne and Legendre, 1997; Hajisamae *et al.*, 2006; Morris and Ball, 2006). Tropical aquatic ecosystems have high biodiversity but are threatened globally (Polunin, 2008). Songkhla Lake is the largest lagoonal water body of Thailand with a basin area of 8,729 km^2 including 1,017 km^2 of main lake water body. This Lake is one example of a tropical shallow lake facing critical water quality deterioration and loss of fish population (Chesoh and Lim, 2008). The Lake comprises three main ecosystems: fresh water in the upper lake, brackish water in the middle lake and saline water in the lower lake which connects to the Gulf of Thailand. Factors that influence such ecosystems include flow seasonality imposed by monsoonal rains and anthropogenic activity, especially destructive fishing (Welcomme *et al.*, 2006). One of the most serious problems is the use of barrier gear to catch fish during their seasonal or ontogenic offshore migration between the lake and the open sea (Katselis *et al*., 2003). Fishing in Songkhla Lake differs spatially according to resources and conditions in each part of the lake. Various types of fishing gear are used in the Lake, both stationary and mobile. Most of these are traps, set-bag nets, and gill nets and such activities affect fish stock dynamics and ecosystems.

Protecting the biodiversity of such lake systems is the challenge faced by ecosystem scientists, who depend on reliable information to guide the protection and utilization of the resources. Important determinants of the quantity of fish catch are gear type (fishing technology) and season. However, few studies (for example, Mahévas *et al*., 2008) have investigated these predictors of fish clustering, and fisheries landing data are commonly subject to errors of various sources resulting in outliers when modeling (Chen *et al*., 2008). In this study, we apply a widely used regression model to cluster fish community structures in fish catch data from Songkhla Lake over a four year period.

The advantage of this study is that it provides a statistically valid model for measurement of species clustering that takes account of how the species are distributed with respect to the season of the year and the type of gear used. It can then be used to provide information for ecosystem–based approaches to fisheries management in Songkhla Lake, and in other shallow tropical habitats.

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 authors. Data include daily records of species, weight of the catch and the three main gear types used (set bag net, trap and gill net).

The set bag net is a traditional fishing gear widely used in the coastal areas of Asia. It is usually placed in a relatively deep channel with a relatively strong current. Having a small mesh size, it is non-selective for size and species. There are currently around 2,200 units, mostly located in the middle and lower zones of the Lake.

The trap is a small-sized stationary fishing gear, combined with seine wing-like barriers used for re-directing and then trapping shrimp and fish. Trap selectivity relies on the fish moving actively into the trap and catches depend on the duration of the soak, i.e. the time the trap is left on the fishing ground. The

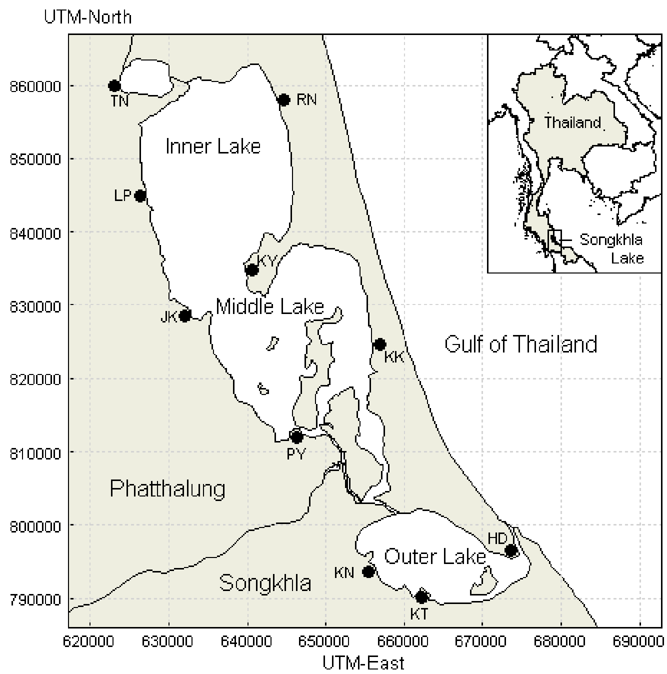

Figure 1. Map of Songkhla Lake and ten fish catch landing sites for data collection.

harvest per unit is rather low but the number of traps is high, currently around 31,200 units, mostly located in the Lower Zone of the Lake (Chesoh and Lim, 2008).

The gill nets are usually long rectangular nets set vertically, where the upper edge has floats while the foot rope has sinkers. Gill nets (both surface net and drift net types) are passive traps that the fish have to swim into to get caught.

There were 127 aquatic animal species found in this study. Swamp eel was excluded from data analysis because it cannot be caught by gill net. Six seasonal bi-month fishing periods were created, namely January-February, March-April, etc. The three types of fishing gear were combined with the six bi-month periods to produce 18 season-gear combinations. The fish landing site was not taken account in the model because the data collected at the landing sites did not accurately identify the location where the fish were caught.

Statistical Analysis

The catch weights used in this paper were computed as totals aggregated over all fishery landings, by species categorized following Choonhapran (1996), and by type of gear and period. Preliminary analysis involved comparing catch weight percentages with respect to species, type of gear used and year for various species groups.

Various statistical methods are available for clustering aquatic and marine organisms according to their patterns of variation in space and time (Hawkins *et al*., 2000; Joy and Death, 2000; Frédou *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 and principal component analyses of covariance matrices.

To remove skewness season-gear comparisons of fish abundance in terms of bi-monthly catch weights were transformed by taking natural logarithms. Since the catch weight (wt) for a given species with a specified gear could be zero, we used the transformation:

$$y = \ln(wt + C), \qquad (1)$$

where C is an appropriate constant. We then fitted a multiple linear regression model to these transformed bi-monthly catches y_{sjt}, where s denotes species with the values 1, 2, ..., S, j denotes season-gear combination with the values 1, 2, …, J and t denotes year with values 1, 2, … T, respectively. The model we used is identical to that used in population science by Booth *et al*. (2002) for forecasting human mortality in terms of age-group and year, and takes the form

$$y_{sjt} = \mu_s + \sum_{k=1}^{m} \alpha_s^{(k)} \beta_j^{(k)} + \varepsilon_{sjt} \, . \qquad (2)$$

This model is an extension of the widely used method proposed by Lee and Carter (1992), which arises in the special case when m is 1. For fixed m, the model can be fitted by least-squares, where the □ components arise as the m eigenvectors of the data covariance matrix corresponding to its largest m eigenvalues.

The model has taxa-specific parameters μ_s encapsulating the variation in catch weights between species, and sets of coefficients $\alpha_s^{(k)}$, k = 1, 2, …, m, denoting the extent to which the taxa have each of the m specific season-gear patterns. Standard errors for these parameters can be obtained approximately by refitting model (2) as a multiple linear regression model in which the $\beta_t^{(k)}$ terms are regarded as fixed predictors. The accuracy of this approximation will depend on the relative numbers of parameters in these predictors compared to the number of data records. In the present application, we found that choosing m=3 gave a satisfactory fit. For this choice, the total number of parameters in the model is 552 (126(m+1) = 504 in μ_s and $\alpha_s^{(k)}$, and 48 in $\beta_t^{(k)}$ allowing for the fact that the m eigenvectors form an orthonormal set), The number of data records is 126x18x4 = 9072 corresponding to the 126 species, 18 gear-season combinations and 4 years, so the number of data records exceeds the number of $\beta_t^{(k)}$ by a factor of 189, so the approximation should be sufficiently accurate for all practical purposes.

A detailed analysis of the goodness-of-fit that highlights individual anomalies involves graphing residuals against corresponding quantiles from the standardized normal distribution.

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

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). They 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%), as shown in Table 1. A total of 127 species belonging to 68 families were caught: 53 marine vertebrate; 30 estuarine vertebrate; 21 freshwater vertebrate; 4

estuarine invertebrate; 18 marine invertebrate and 1 freshwater invertebrate species. The ten families with the highest percentages of catch-weight constituted 62.8% of the total annual landing: 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 four leading species (Broadhead anchovy, Sumatran silverside, Chacunda gizzard shad, and Black tiger shrimp) accounted for 21.5% of the total catch weight, contributing 6.4, 5.3, 5.3 and 4.5 percent, respectively.

Distribution of Catch Patterns

Average catch weights for 126 aquatic animal species (excluding swamp eel which cannot be caught by gill net) using bi-month data groupings for the period 2003 to 2006, were modeled using Equations (1) and (2) with C=0.1 and m=3.

The goodness of fit of the model shows that the model with three components fitted the data reasonably well, with r^2 equal to 0.923. The plot of observed versus fitted values (left panel of Figure 2) shows no evidence of variance inhomogeneity. While the residuals plot indicates some departure from the normality assumption (Figure 2) that could possibly be reduced by increasing the number of components, any such statistical gain needs to be weighed against

Table 1. The 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
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	29.0	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
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
Total	35.8 (1.4%)	277.5 (11.1%)	380.7 (15.2%)	283.9 (11.4%)	849.9 (34.0%)	672.0 (26.9%)	2,499.9 (100%)

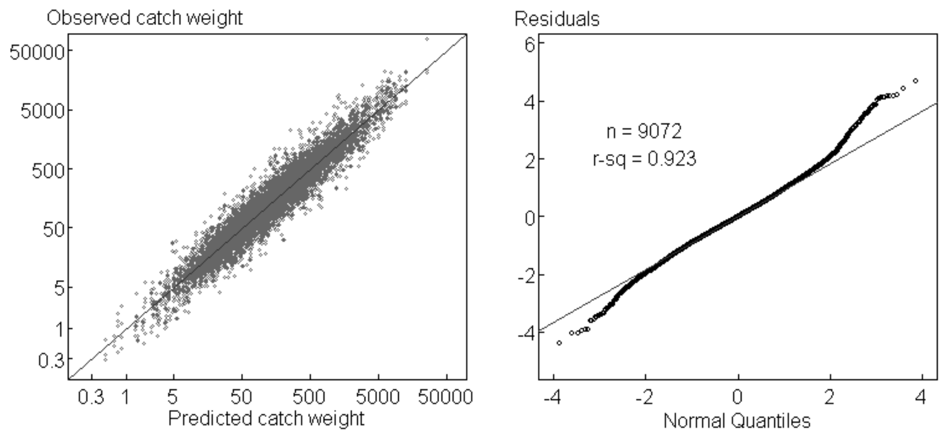

Figure 2. Plots of observed against predicted catch weight (left panel), and residuals against corresponding normal quantiles (right panel).

increased complexity of the model, and with a sample size of 9072 records even a slight departure from normality will be statistically significant. Because the design is balanced the standard errors for the $\alpha_s^{(k)}$ coefficients are all the same, having an estimated value of 0.234. (Note that the μ coefficients reflect the overall catch weights regardless of seasonal changes, and thus depend largely on the biomass of the various species in the Lake.)

The three components ($\beta_t^{(1)}$, $\beta_t^{(2)}$, $\beta_t^{(3)}$) are plotted in Figure 3. All series display interpretable patterns. The first component for each gear type confirms the dominance of the set bag net catch with a maximum catch in the March-April period and some gradual seasonal patterns, peaking in May and June, for the other gear types. Typical species that follow this pattern are characterized by the candy-stripe cardinal fish *(Apogon endekataenia)*, Burmese river gizzard shad, *(Anodontostoma chacunda)*, and streaked spinefoot (*Siganus javus*). The brownback trevally (*Carangoides praeustus*) is representative of the first component and its typical pattern is shown in the upper panel of Figure 4.

The first component thus contained most species of estuarine and marine vertebrates as well as some invertebrates and reflects the fact that set bag net was the gear that resulted in the highest fish catches for these species.

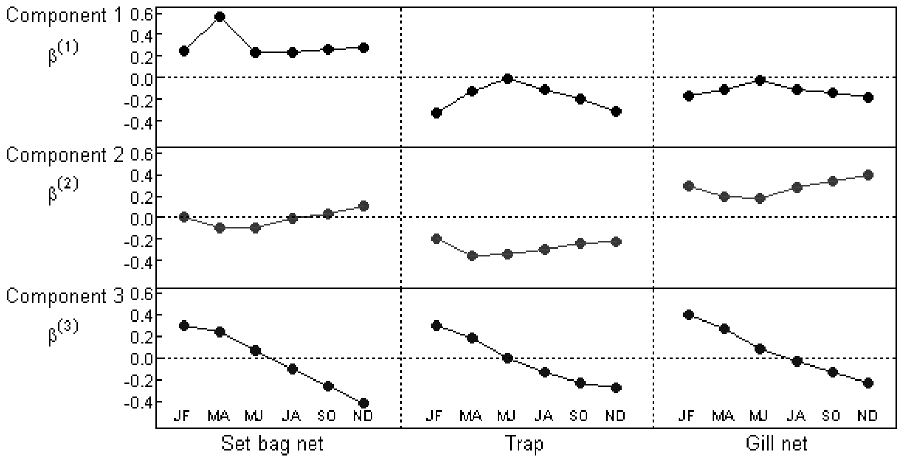

Figure 3. Plots of season-gear patterns for fish catches during 2003-2006.

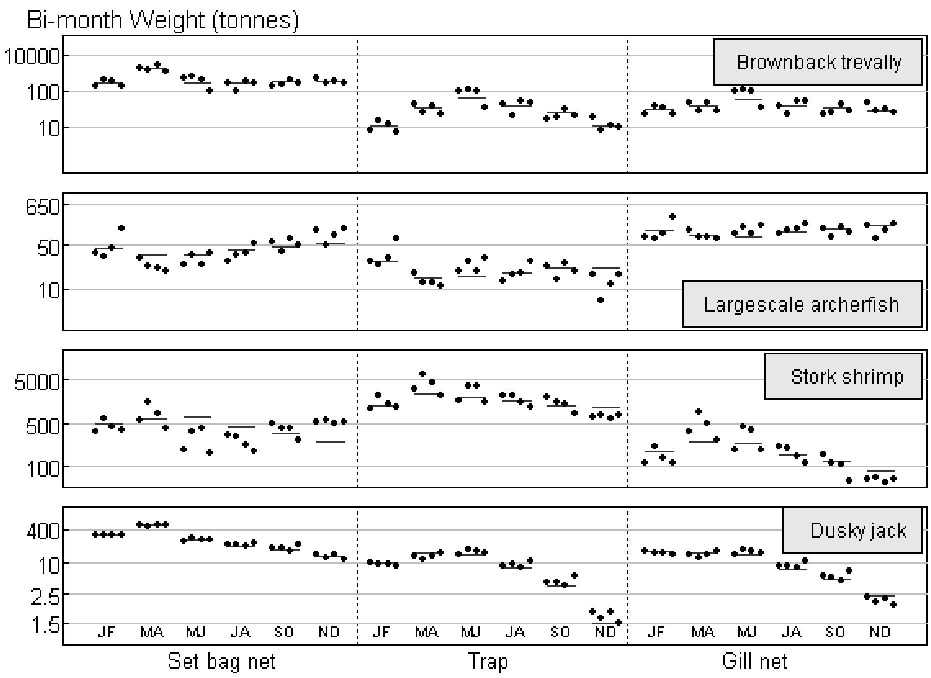

Figure 4. Plots of bi-monthly catch weights in log scale for typical species by each gear type following the dominant patterns from 2003 to 2006.

In contrast, the second component shows a similar seasonal pattern for each gear type with lowest catches in the March to June period, for which the gill net predominates and the trap contributes less. This pattern is characterized by species such as the Gunther walking catfish, (*Clarias macrocephalus*), Giant snakehead (*Channa micropeltes)* and Banana Prawn (*Penaeus merguiensis*). The largescale archerfish (*Toxotes chatareus*), and the stork shrimp (*Metapenaeus tenuipes*) are representative of the second component and its typical pattern is shown in the middle panel of Figure 4. Note that freshwater species were largely caught in gill nets while invertebrates were largely caught in traps.

Finally, the third component has approximately the same magnitude and seasonal pattern for each type of gear, decreasing substantially from highest catches in January-February to lowest catches in November- December. This component shows a seasonal pattern not present in any species: all occurrences were hybrid patterns. The species closest to this pattern was the Dusky jack (*Caranx sexfasciatus*) as shown in the lower panel of Figure 4.

Clustering of Fish Community Structure

Figure 5 shows various views of the estimated coefficients for the three components. Distinctive fish community clusters can be seen clearly. The dominant cluster comprises 90 species containing all marine and estuarine vertebrate species and seven species of marine invertebrates. Two distinct sub-clusters are clearly separated, as well as a small group of four species. The first sub-cluster of 20 species comprises all freshwater vertebrates, and the second comprises 12 marine invertebrate species. The small group contains the Indopacific mackerel (*Scomberomorus guttatus*), streaked seerfish (*Scomberomorus lineolatus*) and bigeye trevally (*Caranx sexfasciatus*), as well as an isolated species of cuttlefish (*Sepioteuthis lessoniana*).

Discussion

Throughout the study, our statistical model identified three component patterns for species catch weights. The model provided an r-squared value of 92.3%. With the first three components, 93% of the total variance could be explained. The model was very effective in clearly separating fish community clusters with absolutely no overlap between clusters, freshwater and saltwater fish. The patterns for each typical component showed that bi-monthly fluctuations in the abundance of fish species assemblages depended on the fishing gear used, especially for the first component, and for the second

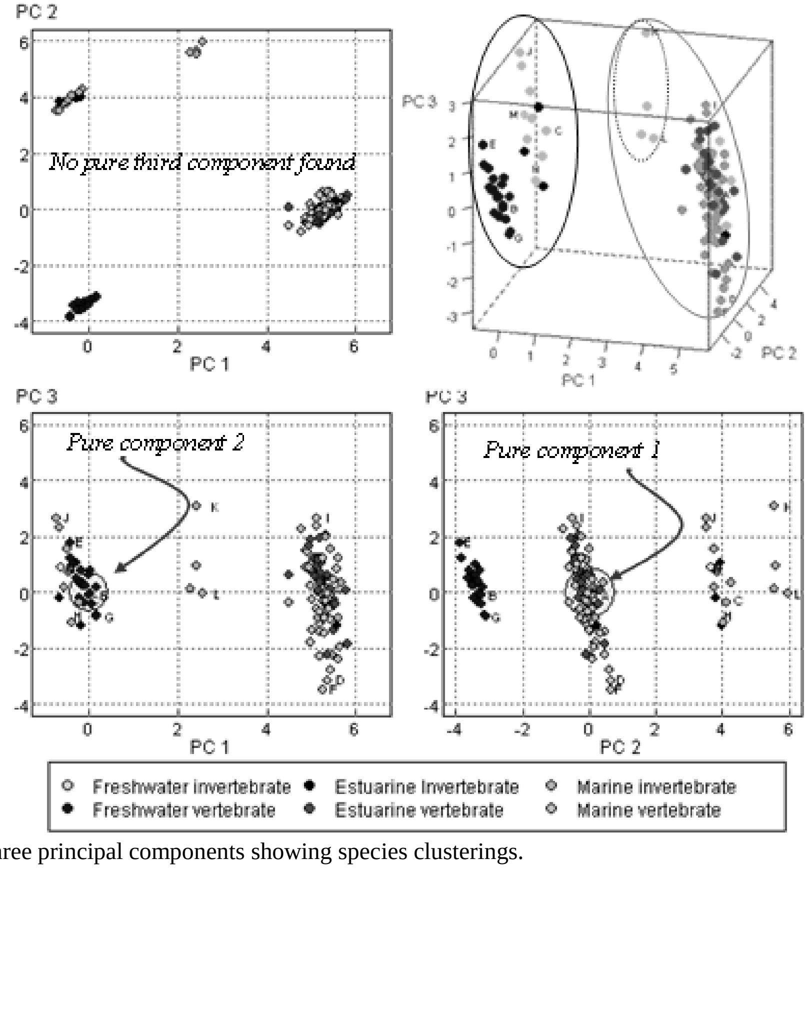

Figure 5. Plots of three principal components showing species clusterings.

component depended on the interaction between season and type of gear. The third component pattern focused on the seasonal fluctuation in catch weight, not exactly followed by any species found.

Fish diversity, recruitment and production depend greatly on limnological characteristics. Fish species composition in a tropical shallow lake, like Songkhla Lake, requires a species-specific environment. Generally, fluctuations of the water quality affect the species distribution, with aquatic animal species either tolerant to the environment or migrating to another environment (Katselis *et al*., 2003; Labropoulou and Papaconstantinou, 2004). Fishing is usually based on trapping the fish on their passage from their feeding ground to spawning and nursery grounds. Traditionally, fish barrier traps are stationary passive gear and their catch depends on their physical features and placement location. Catch landings also depend on the type of gear used. Gill nets have high catchability for freshwater fish because they are the most common gear used in the upper lake zone, where the water is always turbid and windswept, reducing visibility for freshwater fish and so increasing catchability. In contrast, the giant freshwater prawn (*Macrobrachium rosenbergii*), and some marine shrimps and crabs comprise the highest catch in traps.

The patterns also indicated increasing freshwater catch weights, while marine invertebrate catches decreased, signaling that lake fisheries resources need to be regulated. Fishing is the most important human impact affecting the fish community structure in lakes (Kangur *et al*., 2007). Therefore, restricting the number and placement of traps will increase efficiently the stocks of all economic invertebrates, such as the blue swimming crab, acetes, cross-marked swimming crab, serrated swimming crab, squid, cuttlefish, octopus and giant freshwater prawn. Similar control of gill nets will be effective in increasing populations of all freshwater fish, and controlling set bag nets should increase stocks of estuarine and marine vertebrates.

Although several species are ubiquitous in all habitats, others are anadromous, catadromous, amphidromous, and oceanodromous and so have specific habitats in Songkhla Lake. Our regression model, based on the log-transformed catch weights of classified species and season-gear combinations, clearly separates distinctive fish community clusters.

The overall fishing effort in the Lake was reasonably stable over the period of the study, with both fishing gear and number of fishermen not changing substantially (NICA, 2007). While the study had some limitations, in that fishing effort and correlation of the environmental parameters were not taken into account, we recommend that this model can provide more interpretable information and will generally benefit the understanding for clustering of fish community structure in fishery data.

Acknowledgements

We would like to express our sincere gratitude to Prof Don McNeil, Dr Apiradee Lim, Dr Sukree Hajisamae, and Dr Phattrawan Tongkumchum for their invaluable assistance, encouragement and helpful guidance. We also thank the Songkhla Provincial Fisheries Office, the Phattalung Provincial Fisheries Office and the National Institute of Coastal Aquaculture (NICA), Department of Fisheries, Thailand, for allowing us to use their data.

References

Booth, H., Maindonard, J. and Smith, L. 2002. Applying Lee-Carter under conditions of variable mortality decline. Population Studies, 56: 325-336.

Chen, Y., Chen, X. and Xu, L. 2008. Developing a size indicator for fish populations. Scientia Marina, 72: 221–229.

Chesoh, S. and Lim, A. 2008. Forecasting fish catches in the Songkhla Lake basin. ScienceAsia, 34: 335–340.

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, 32 pp.

Clarke, K.R. and Warwick, R.M. 1994. Change in marine communities: An approach to statistical analysis and interpretation. PRIMERE Ltd, Plymouth, 144 pp.

Dufrêne, M. and Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs, 67: 345–366.

Frédou, T., Ferreira, B.P. and Letourneur, Y. 2006. A univariate and multivariate study of reef fisheries off northeastern Brazil. ICES Journal of Marine Science, 63: 883–896.

Hajisamae, S., Yeesin, P. and Chaimongkol, S. 2006. Habitat utilization by fishes in a shallow, semi−enclosed estuarine bay in southern Gulf of Thailand. Estuarine, Coastal Shelf Science, 68: 647–655.

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

Joy, M.K. and 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 and Freshwater Research, 34: 243–254.

Kangur, K., Park, Y.-S., Kangur A., Kangur, P. and Lek, S. 2007. Patterning long-term changes of fish community in large shallow Lake Peipsi. Ecological Modelling, 203: 34-44.

Katselis, G., Koutsikopoulos, C., Dimitriou, E. and 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.

Labropoulou, M. and Papaconstantinou, C. 2004. Community structure and diversity of demersal fish assemblages: the role of fishery. Scientia Marina, 68: 215–226.

Lee, R.D. and Carter, L.R. 1992. Modelling and forecasting U.S. mortality. Journal of the American Statistical Association, 87: 659–671.

Mahévas, S., Bellanger, L. and Trenkel, V.M. 2008. Cluster analysis of linear model coefficients under contiguity constraints for identifying spatial and temporal fishing effort patterns. Fisheries Ressearch, 93: 29–38

Morris, L. and Ball, D. 2006. Habitat suitability modelling of economically important fish species with commercial fisheries data. ICES Journal of Marine Science, 63: 1590–1603.

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

Polunin, N.V.C. 2008. Aquatic ecosystems: trends and global prospects. Cambridge University Press, Cambridge, 499 pp.

Venables, W.N. and Smith, D.M. 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. (May 20, 2007)

Welcomme, R.L., Winemiller, K.O. and Cowx, I.G. 2006. Fish environmental guilds as a tool for assessment of ecological condition of rivers. River Research Applications, 22: 377–396.