

Appendix 2

**Publication 2 (Manuscript): Regression-based modeling of
macrobenthic fauna density in the Middle Songkhla Lake, Thailand**

*Prince of Songkhla University
Pattani Campus*

Regression-Based Modeling of Macrobenthic Fauna Density in Middle Songkhla Lake, Thailand

Uraiwan Sampantarak*[a], Noodchanath Kongchouy [b], and Saowapa Angsupanich [c]

[a] Pattani Inland Fisheries Research and Development Center, Department of Fisheries, Pattani 94160, Thailand

[b] Department of Mathematics, Faculty of Science, Prince of Songkla University, Songkhla 90112, Thailand

[c] Department of Aquatic Science, Faculty of Natural Resources, Prince of Songkla University, Songkhla 90112, Thailand

* Author for correspondence; e-mail: uraiwan111@hotmail.com, Tel: +66 8 4653 9391

ABSTRACT

This study examined distributional patterns of macrobenthic fauna assemblages in relation to environmental characteristics in the middle of Songkhla Lake, Thailand. Macrobenthic fauna and water quality parameters including sediment characteristics were obtained from nine sampling sites at bimonthly intervals from April 1998 to February 1999. Factor analysis was used to define five predictors including three composite variables based on salinity, physical sediment characteristics, and physico-chemical properties of water and sediment, together with total suspended solids and dissolved oxygen as single variables. A multivariate multiple regression model (MMR) was used to examine relationships between these predictors and the densities of twenty four selected macrobenthic families with greater than 35% occurrence. To remove skewness, the densities were log-transformed before fitting the model. Results were compared with those obtained using canonical correspondence analysis. MMR can be used as additional or alternative method to analyse relationship between environmental variables and abundance of benthic organisms in coastal ecosystem.

Keywords: Macrobenthic fauna, Multivariate multiple regression model, Factor analysis, Canonical correspondence analysis, Tropical lagoon

1. INTRODUCTION

Macrobenthic fauna are recognized as sensitive indicators of environmental disturbance (Weisberg et al. [1], Borja et al. [2], Ranasinghe et al. [3]). They have limited mobility. Many of them are unable to avoid adverse conditions brought about by natural stresses or human impacts. Moreover, their relative longevity, with many species having life spans in excess of two years, allows them to integrate responses to environmental processes over extended time periods (Gray et al. [4]). In addition, observed distribution of macrobenthic fauna are useful in diagnostic studies and environmental monitoring (Warwick [5]).

Clarke and Warwick [6] outlined the basic methods now commonly used by biological scientists for analysis of their data. For descriptive studies these methods include data transformation using square roots, fourth roots or logarithms (after adding 1 to cell counts or densities to handle zeros) to remove skewness, principal components analysis of covariance matrices, and ordination procedures to cluster taxa in space and time, as well as more complex multivariate analytical techniques such as dendrograms based on similarity matrices and multidimensional scaling. Measures of association in assemblage data such as the Bray-Curtis similarity index are preferred to Pearson correlation coefficients “for sound biological reasons” (Clarke et al. [7]), but such measures do not satisfy the positive-definiteness assumptions that underpin conventional multivariate statistical analysis.

For comparative studies to assess associations between species abundance outcomes and environmental predictor variables, canonical correspondence analysis (Ter Braak [8]) is now used

extensively in the biological literature (von Wehrden et al. [9]). Some important studies using this method include those reported by Rakocinski et al. [10], Hawkins et al. [11], Joy and Death [12], Guerra-García et al. [13], Hajisamae and Chou [14], Morrissey et al. [15], Ysebaert et al. [16], Quintino et al. [17], Anderson [18], Glockzin and Zettler [19].

Although exceptions exist such as studies by Liang et al. [20] using structural equation modeling and by Warton and Hudson [21] using multivariate analysis of variance, multivariate multiple regression analysis is not commonly used in the biological literature for analyzing species abundance patterns. However, this method would appear to be an ideal statistical method for assessing relationships between species abundance outcomes and their environmental predictors, for the simple reason that it is the natural extension of ordinary regression analysis involving a single outcome to any number of mutually correlated outcomes such as species abundances. It is thus of interest to compare this method with its biologically preferred counterpart using common sets of biological data relating taxonomic abundances to environmental determinants, and this is the object of our study.

For this comparison we used data from a study involving macrobenthic fauna abundances and various water and sediment characteristics collected at specified locations in an estuarine lake over a period of one year reported by Angsupanich et al. [22]. The methods compared are canonical correspondence analysis (CCA) using CANOCO Version 4.5 (Ter Braak and Šmilauer [23]) and multivariate multiple regression (MMR) using R Version 2.10.0 (R Development Core Team [24]).

2. MATERIALS

Songkhla Lake is a shallow coastal lagoon, located in a tropical coastal ecosystem in Southern Thailand. It covers an area of 1,040 km² with 20 km width and 75 km length, approximately. It is divided into three parts as the Upper Lake in Phatthalung Province, the Middle Lake between borders of Songkhla Province and Phatthalung Province, and the Lower Lake in Songkhla Province connected to the Gulf of Thailand. Some canals pour fresh water into the lake. The salinity slowly increases where the freshwater and seawater meet. Thus, the water in the Middle Lake is brackish, and becomes saltier in the area around the lake mouth (Lower Lake). The zone of interest for this study covers an area of 390 km² located between UTM 635000E and 660000E in the west-east direction and between UTM 840000N and 805000N in the north-south direction (Figure 1).

Angsupanich et al. [22] collected macrobenthic fauna using a Tamura's grab (0.05 m²) from the nine sampling stations. The assemblages were conducted with 11 replications for each station at bimonthly intervals from April 1998 to February 1999. The samples were sieved consecutively through the screens and fixed in 10% Rose Bengal-formalin for later identification.

The densities of macrobenthic fauna were recorded as the number of individuals per square meter (ind m⁻²) for each species. A total of 161 taxa of macrobenthic fauna were found and classified into 81 families. In many cases the species could not be identified exactly, so in our model the outcomes were classified by family instead of species. With nine locations and six bimonthly data study periods, we defined the occurrence for a specified family as the proportion of these 54 occasions on which at least one organism was found. We then selected the 24 families with greater than 35% occurrence (93.2% total assemblages) for data analysis.

Environmental variables comprised water depth (wDep), water temperature (wTemp), salinity (Sal), water pH (wpH), dissolved oxygen (DO), total suspended solids (TSS), with sediment pH (spH), total nitrogen content (TN), organic carbon content (OC), and soil structure (percentages of sand, silt, and clay). These were measured with three replications on the same occasions as the biotic data.

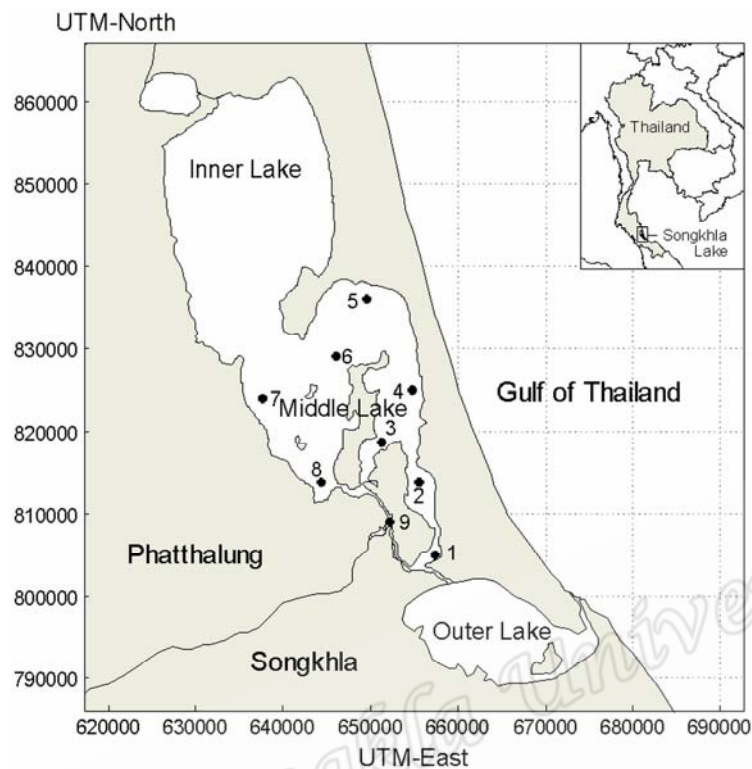


Figure 1. Songkhla Lake and sampling sites (labeled 1-9)

3. METHODS

The nine sampling stations and six bimonthly periods were combined as 54 station-month measurement occasions. The response variable was taken as $\log(1 + c \times \text{density})$ with the multiplier c chosen to approximate normality of error distributions. The predictors comprised environmental components derived from a factor analysis together with unique variables not accommodated by the factor analysis.

FACTOR ANALYSIS

Factor analysis is performed on the environmental variables with the aim of substantially reducing correlations between them that could mask their associations with the outcome variables. Each factor identifies correlated groups of variables. Ideally each group (which must contain at least two variables to contribute to the factor analysis) contains variables having small correlations with variables in other groups. To achieve this, any variable uncorrelated with all other variables is omitted from the factor analysis. Each factor comprises weighted linear combinations of the variables and these factors are rotated to maximize the weights of variables within the factor group and minimize the weights of variables outside the group. The resulting weights are called “loadings”. Variables omitted from the factor analysis due to low correlation with all other variables (high “uniqueness”) are treated as separate predictors, so predictors include single variables as well as factors.

The number of factors selected was based on obtaining an acceptable statistical fit using the chi-squared test, and these factors were fitted using maximum likelihood with promax rotation in preference to varimax, which requires the rotation to be orthogonal (Browne [25], Abdi [26]).

MULTIVARIATE MULTIPLE REGRESSION

Multivariate multiple regression (MMR) is used to evaluate the effects of multiple predictor variables on multiple response variables. The model (Mardia et al. [27]) may be defined in matrix form as

$$\mathbf{Y}_{(n \times p)} = \mathbf{X}_{(n \times q)} \mathbf{B}_{(q \times p)} + \mathbf{E}_{(n \times p)}. \quad (1)$$

In this formulation $\mathbf{Y}_{(n \times p)}$ is an observed matrix of p response variables on each of n occasions, $\mathbf{X}_{(n \times q)}$ is the matrix of q predictors (including a vector of 1s) in columns and n occasions in rows, $\mathbf{B}_{(q \times p)}$ contains the regression coefficients (including the intercept terms), and $\mathbf{E}_{(n \times p)}$ is a matrix of unobserved random errors with mean zero and common covariance matrix Σ . Ordinary (univariate) multiple regression arises as the special case when $p = 1$. If $q - 1$ environmental predictors $f_i^{(k)}$ ($k = 1, 2, \dots, q - 1$) are available, the prediction model for outcome j on occasion i may be expressed as

$$y_{ij} = \mu_j + \sum_{k=1}^{q-1} \beta_j^{(k)} f_i^{(k)}. \quad (2)$$

The model fit may be assessed by plotting the residuals against normal quantiles (Venables and Ripley [28]), and also by using the set of r-squared values for the response variables to see how much of the variation in each is accounted for by the model.

The method also provides standard errors for each of the $p \times q$ regression coefficients thus providing p -values for testing their statistical significance after appropriate allowance for multiple hypothesis testing. The multivariate analysis of variance (MANOVA) decomposition is also used to assess the overall association between each environmental predictor and the set of outcomes by the likelihood ratio, Pillai's trace criterion (Olson [29], Johnson and Wichern [30]).

CANONICAL CORRESPONDENCE ANALYSIS

Assuming that the data structure comprises the \mathbf{Y} and \mathbf{X} matrices with rows corresponding to measurements of outcomes and predictors taken on the same occasions, canonical correspondence analysis (Ter Braak [8]) produces a two-dimensional *biplot* comprising arrows of variable lengths and directions (*gradients*) emanating from a common origin representing the predictor variables, together with superimposed points denoting the outcome variables. The relative lengths of the arrows and the angles between them are based on the correlation matrix of the predictor variables, and the coordinates of the points are planar projections of the density outcomes, computed in such a way that their positions relative to the arrows portray their associations with the environmental predictors. The method also produces coordinate scores and p -values for the overall associations based on Monte Carlo permutation tests.

4. RESULTS

ENVIRONMENTAL PARAMETERS

Figure 2 plots the water characteristics in Middle Songkhla Lake from April 1998 to February 1999. The water depth varied to a lesser extent, but was also higher during the rainy season, varying with location from an average of less than 1 m at stations four and nine to more than 2 m at station eight. The water temperature showed decreased values in the rainy season, with range 27-34°C. The salinity increased from close to zero during the rainy season (December to February) to an average close to 20 in other months. The pH of water was also lowest in December.

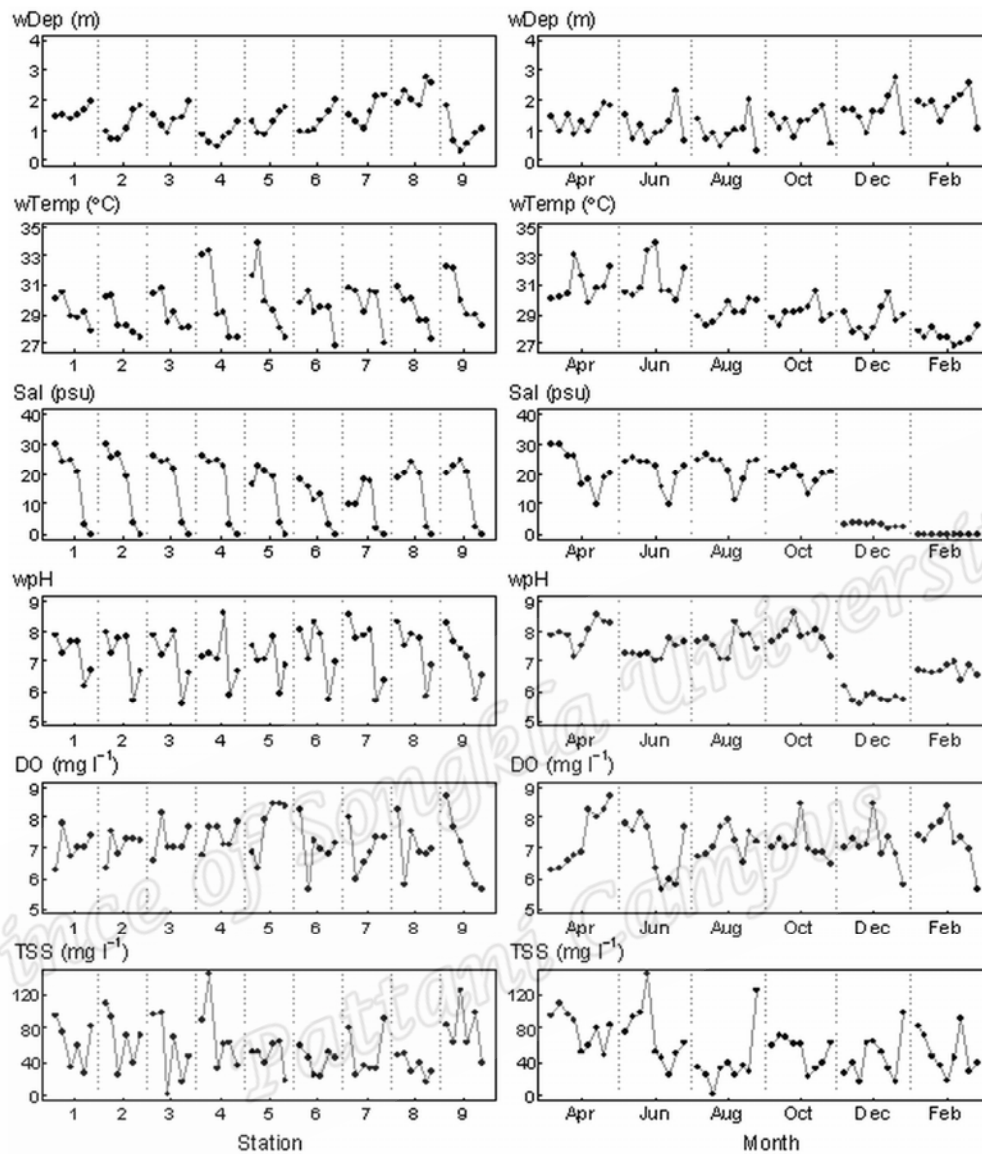


Figure 2. Water characteristics in Middle Songkhla Lake from April 1998 to February 1999 by station (left panel) and month (right panel)

Figure 3 plots sediment characteristics measured on the same occasions as the water characteristics. The total nitrogen content at each station was very low (0.02%) from October to February. The organic carbon content was relatively constant with respect to month, but showed the highest value at station nine in every month except August. The lake bed at station six was mostly characterized by sand (mean = 84.6%) and station 9 was mostly characterized by clay (mean = 53.2%), also with high values of organic carbon. Note that the sand, silt, and clay percentages sum to 100%.

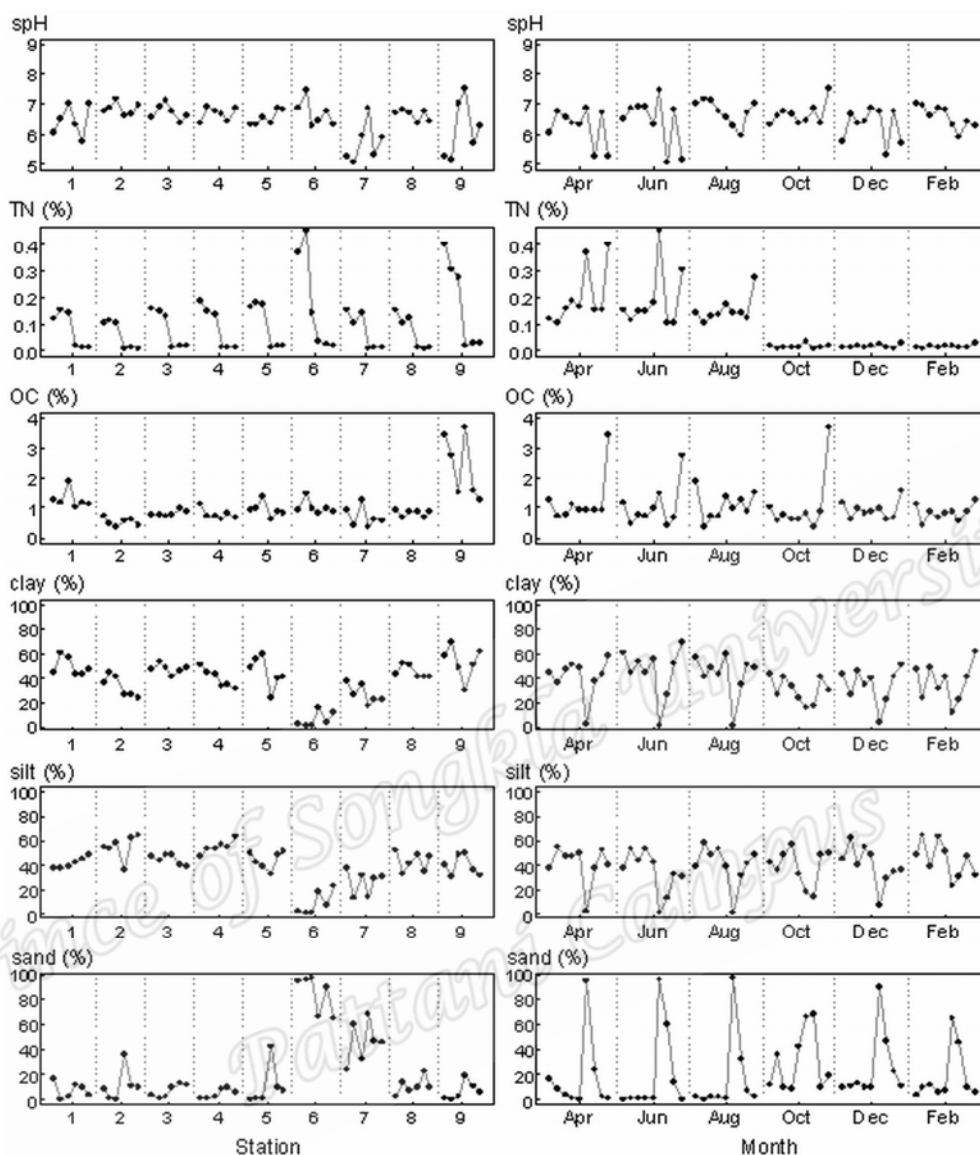


Figure 3. Sediment characteristics in Middle Songkhla Lake from April 1998 to February 1999 by station (left panel) and month (right panel)

OCCURRENCE AND ABUNDANCE OF MACROBENTHIC FAUNA

Table 1 shows the taxa percentages of occurrence and density in individuals per square meter of the 24 families of macrobenthic fauna measured with the water characteristics. A total of 24 families were classified in three phyla of Annelida (Polychaeta), Arthropoda (Crustacea) and Mollusca (Gastropoda and Bivalvia), which comprised the most diverse group (35.2-98.2% of occurrence). Polychaeta was represented by nine families (Capitellidae, Goniadidae, Hesionidae, Nephtyidae, Nereididae, Pilargüidae, Pholoidae, Spionidae and Terebellidae). Crustacea was also represented by nine families (Aoridae, Isaeidae, Melitidae, Oedicerotidae, Apseudidae, Pseudotanaididae, Anthuridae, Cirolanidae and Alpheidae). Marginellidae, Retusidae, Skeneopsidae and Stenothyridae were in the Gastropoda whilst the two remaining families (Tellinidae and unidentified species were in the Bivalvia).

Table 1. Taxa occurrence (%occ) and density in individual per square meter (ind m⁻²) of 24 families of macrobenthic fauna in Middle Songkhla Lake from April 1998 to February 1999. The asterisk is the unidentified species in Bivalvia

Phylum	Class	Order	Family	%occ	Density (ind m ⁻²)
Annelida	Polychaeta	Capitellida	Capitellidae (Cap)	87.0	1,227.3
	Polychaeta	Phyllodocida	Goniadidae (Gon)	37.0	443.6
	Polychaeta	Phyllodocida	Hesionidae (Hes)	55.6	698.2
	Polychaeta	Phyllodocida	Nephtyidae (Nep)	87.0	2,218.2
	Polychaeta	Phyllodocida	Nereididae (Ner)	98.2	8,507.3
	Polychaeta	Phyllodocida	Pilargiidae (Pil)	70.4	1,625.5
	Polychaeta	Phyllodocida	Pholoidae (Pho)	59.3	658.2
	Polychaeta	Spionida	Spionidae (Spi)	92.6	5,056.4
	Polychaeta	Terebellida	Terebellidae (Ter)	35.2	1,136.4
Arthropoda	Crustacea	Amphipoda	Aoridae (Aor)	59.3	2,421.8
	Crustacea	Amphipoda	Isacidae (Isa)	87.0	6,900.0
	Crustacea	Amphipoda	Melitidae (Mel)	94.4	4,438.2
	Crustacea	Amphipoda	Oedicerotidae (Oed)	66.7	667.3
	Crustacea	Tanaidacea	Apseudidae (Aps)	90.7	40,083.6
	Crustacea	Tanaidacea	Pseudotanaididae (Pse)	37.0	4,265.5
	Crustacea	Isopoda	Anthuridae (Ant)	75.9	3,816.4
	Crustacea	Isopoda	Cirolanidae (Cir)	37.0	427.3
	Crustacea	Decapoda	Alpheidae (Alp)	40.7	98.2
Mollusca	Gastropoda	Neogastropoda	Marginellidae (Mar)	85.2	3,963.6
	Gastropoda	Cephalaspidea	Retusidae (Ret)	55.6	5,536.4
	Gastropoda	Mesogastropoda	Skeneopsidae (Ske)	38.9	956.4
	Gastropoda	Mesogastropoda	Stenothyridae (Ste)	35.2	581.8
	Bivalvia	Unidentified	Unidentified* (Uni)	44.4	338.2
Bivalvia	Veneroidea	Tellinidae (Tel)	81.5	17,134.5	

Nereididae was the most commonly observed family with 98.2% occurrence whereas Terebellidae and Stenothyridae had lowest occurrence (35.2%). Apseudidae was the most abundant family with average density of 40,083.6 ind m⁻²; on the other hand, the Alpheidae was least abundant with average density 98.2 ind m⁻².

FACTOR ANALYSIS

DO and TSS were omitted from the factor analysis due to high uniquenesses (0.975 and 0.848, respectively). The model provided an adequate fit using three factors (chi-squared = 20.23, 12 df, *p*-value = 0.063). Table 2 shows the loadings, with values less than 0.20 in magnitude suppressed. If only loadings greater in magnitude than 0.45 are considered, the three factors do not contain any overlapping variables.

Factor 1 encompasses salinity, containing positive loadings for Sal and wpH, and a negative loading for wDep as expected, with deeper water during the rainy season. Factor 2 represents the effect of sediment characteristics in the lake bed (sand-clay habitat), consisting of a positive loading for sand and a similar negative loading for clay. Factor 3 characterizes physical and chemical compositions in the lake, comprising positive loadings for TN, OC, and wTemp, and a negative loading for spH. Thus Factor 1 was defined as $-0.53 \times wDep + 0.70 \times wpH + 0.99 \times Sal$, Factor 2 as $0.94 \times Sand - 0.95 \times Clay$, and Factor 3 as $0.47 \times OC + 0.58 \times TN - 0.57 \times spH + 0.54 \times wTemp$. The three factors respectively accounted for 24.6%, 20.7%, and 13.7% of the variance in the environmental data, a total of 59.0%. The three factors were included in the regression model as predictors together with the two singleton variables omitted from the factor analysis, with each of these five predictor variables scaled to have mean 0 and standard deviation 1.

Table 2. Factor analysis (with loadings below 0.2 omitted)

Environmental variables	Factor 1	Factor 2	Factor 3
Organic carbon (OC)	-	-	0.47
Total nitrogen (TN)	0.34	-	0.58
Sediment pH (spH)	0.39	-	- 0.57
Water depth (wDep)	- 0.53	-	-
Water pH (wpH)	0.70	-	-
Salinity (Sal)	0.99	-	-
Water temperature (wTemp)	0.42	-	0.54
Sand	-	0.94	-
Clay	-	- 0.95	-
% Total variance	24.6	20.7	13.7
% Cumulative variance	24.6	45.3	59.0

REGRESSION ANALYSIS

The choice $\epsilon = 100$ in the transformation $\log(1 + \epsilon \times \text{density})$ gave residuals satisfying the normality assumption. The left panel of Table 3 shows the corresponding individual regression coefficients and standard errors and r-squared values for each family after fitting the MMR model with all five environmental predictors included. The right panel shows the corresponding results for a reduced model containing only the two predictors that were statistically significant in the MANOVA (Table 4).

Table 3. Coefficients and standard errors (in parenthesis) from fitting multivariate multiple regression models with all five environmental predictors (left panel) and with only the two predictors found statistically significant in the MANOVA (right panel). Coefficients with p -values greater than 0.05 in both models are omitted; those adjudged not honestly statistically significant are shown in italics and those with p -values less than 0.01 are shown in bold.

Fam	5 predictors					2 predictors			
	Factor 1	Factor 2	Factor 3	-TSS	DO	r ²	Factor 1	Factor 2	r ²
Hes	0.76 (0.28)	-	-	-	-	0.23	0.67 (0.25)	-	0.18
Uni	0.71 (0.27)	-	-	-	-	0.16	0.53 (0.25)	-	0.08
Spi	0.44 (0.21)	-	-	-	-	0.13	0.43 (0.19)	-	0.09
Gon	1.01 (0.26)	-	-0.76 (0.31)	1.08 (0.45)	-	0.31	0.61 (0.25)	-	0.11
Pho	-0.61 (0.25)	0.43 ^{ns} (0.22)	-	-	-	0.33	-0.79 (0.23)	0.47 (0.22)	0.28
Pse	-0.64 (0.28)	0.81 (0.24)	-	-	-	0.38	-0.65 (0.25)	0.85 (0.24)	0.32
Ant	-	0.69 (0.24)	-	-	-	0.19	-	0.73 (0.24)	0.17
Pil	-	-0.68 (0.25)	-	-	-	0.17	-	-0.66 (0.24)	0.16
Aor	-	0.87 (0.26)	-	-1.11 (0.53)	-	0.26	-	0.75 (0.27)	0.14
Ter	-	1.04 (0.20)	0.60 (0.28)	0.92 (0.41)	-	0.45	-	1.08 (0.21)	0.35
Cir	0.42 ^{ns} (0.24)	0.51 (0.21)	0.67 (0.29)	-	-	0.30	0.69 (0.23)	0.42 ^{ns} (0.21)	0.19
Ner	0.27 ^{ns} (0.15)	0.30 (0.14)	0.42 (0.19)	-	-	0.25	0.41 (0.14)	0.27 ^{ns} (0.14)	0.17
Ste	0.35 ^{ns} (0.25)	-	0.65 (0.30)	-	-	0.25	0.52 (0.23)	-	0.17
Mel	0.20 ^{ns} (0.18)	-	0.58 (0.22)	-	-	0.24	0.42 (0.18)	-	0.11
Ret	-	-	-	1.56 (0.58)	-	0.22	-	-	0.05
Cap	-	-	-	-	-0.86 (0.36)	0.15	-	-	0.07
Nep	-	-0.16 ^{ns} (0.24)	-	-	-	0.12	-	-0.42 (0.20)	0.08
Isa	-	0.46 ^{ns} (0.23)	-	-	-	0.16	-	0.47 (0.23)	0.09
Mar	-	0.42 ^{ns} (0.22)	-	-	-	0.16	-	0.45 (0.21)	0.15
Alp	-	-	-	-	-	0.09	-	-	0.04
Oed	-	-	-	-	-	0.08	-	-	0.04
Aps	-	-	-	-	-	0.05	-	-	0.04
Ske	-	-	-	-	-	0.02	-	-	0.01
Tel	-	-	-	-	-	0.10	-	-	0.08

The coefficients listed are the ones statistically significant at 5% and 1% (in bold). Since there are 120 regression coefficients in all and 5% of these would be expected to have p -values

less than 0.05 even if all their corresponding population parameters were zero, the six largest p -values less than 0.05 are italicized to indicate failure to achieve “honest” significance. The additional coefficients (labeled m) were not statistically significant in their fitted model, but achieved significance in the other model. Note that the coefficients for TSS are reversed in sign because most were negative, so this predictor is labeled $-TSS$.

Table 4. MANOVA decomposition for multivariate multiple regression model with five predictors.

Source of variance	Df	Pillai	approx F	Df (num)	Df (denom)	Prob (>F)
Intercept	1	0.990	104.092	24	25	< 0.0001
Factor 1	1	0.793	3.968	24	25	0.0005
Factor 2	1	0.778	3.641	24	25	0.0010
Factor 3	1	0.596	1.535	24	25	0.1468
□ TSS	1	0.503	1.055	24	25	0.4466
DO	1	0.339	0.691	24	25	0.8157
Residuals	48					

CANONICAL CORRESPONDENCE ANALYSIS

Figure 4 shows biplots based on the CCA matching the two MMR analyses. The families are represented by dots whereas the environmental predictors are represented by arrows. Each arrow determines an axis in the plots, obtained by extending the arrows in both directions.

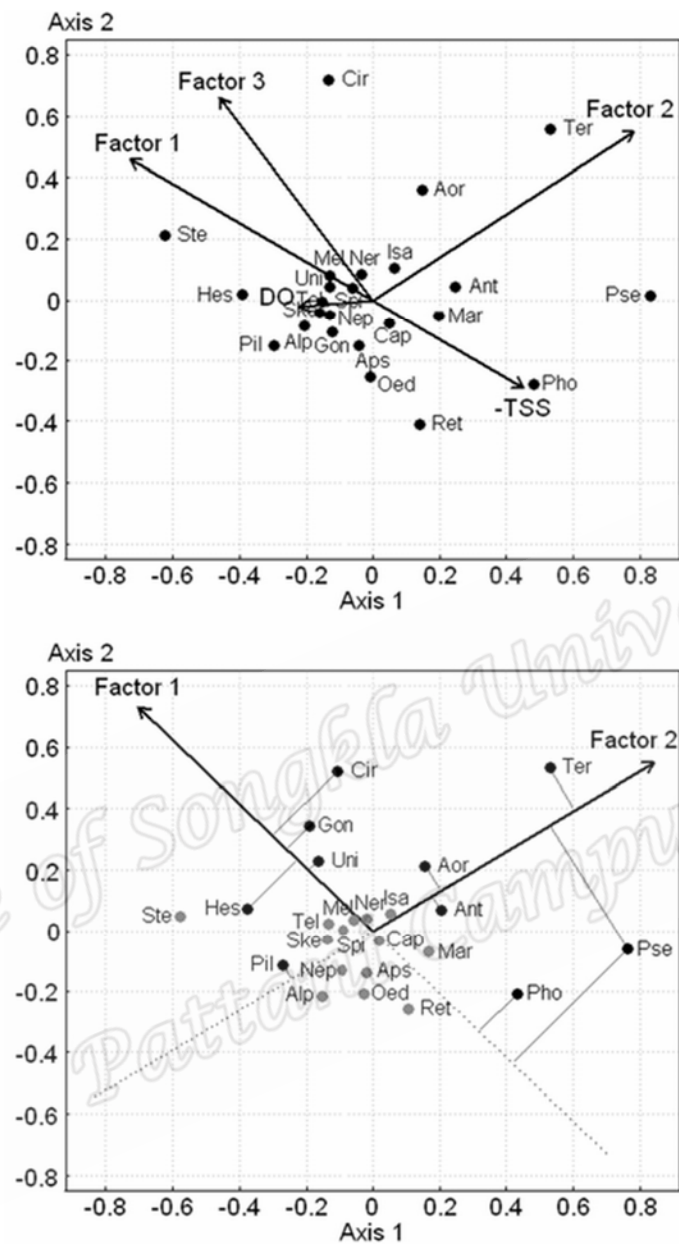


Figure 4. Upper panel: biplot of first two axes of CCA ordination diagram with 24 macrobenthic fauna families against five environmental predictors. Lower panel: similar biplot omitting the three environmental variables not statistically significant in the MANOVA, with families showing highly significant coefficients in the multivariate multiple regression model (p -values < 0.01) connected to the corresponding arrows representing the predictors

5. DISCUSSION

COMPARISON METHODS

This study aim was to identify potential environmental “key factors” accounting for the distribution of macrobenthic fauna communities, to improve understanding concerning benthic/abiotic interactions and ecosystem functioning. By replacing groups of correlated predictors by single variables, factor analysis was used to remove correlations between environmental parameters that mask their effects on the macrobenthos densities.

Multivariate multiple regression and canonical correspondence analysis were used to examine the relations between the macrobenthic fauna family densities and the reduced set of environmental predictors. Although each of these five predictors was found to be associated with at least one family, the corresponding MANOVA decomposition found only two of them to be statistically significant overall. The biplot produced by the canonical correspondence analysis is seen to be more informative when only these two predictors were included.

The MMR model containing all five predictors gives seven associations between a family density and an environmental determinant that are highly statistically significant (p -value < 0.01), and a further nine with p -value between 0.01 and 0.05. Ten families showed no evidence of an association with any of the five determinants. Most of these associations can also be seen in the CCA biplot.

The most noticeable difference between the results of the two methods is that Spionidae is found to be associated with the salinity factor in the MMR model but this association is not seen in the biplots. Since Spionidae is a marine benthos (Day and Blake [31]) with typical dominant species *Pseudopolydora kempfi* and *Prionospio cirrifera* (Angsupanich et al. [22]), there is evidence supporting the MMR result in this case. In the biplot containing all five environmental predictors the arrows for Factor 1 and TSS have identical directions, but the correlation between these variables (0.30) is not high.

SCIENTIFIC FINDINGS

The results, both by MMR and CCA, clearly indicate that the salinity factor was positively associated with the densities of the Goniadidae, Hesionidae, and Spionidae and the unidentified families in the Bivalvia, and negatively associated with the densities of Pholoidae and Pseudotanaidae. This is in contrast with those analyzed based on the same data using BIOENV by Angsupanich et al. [22] which indicated no impact of salinity on benthos density. In general, salinity is an important factor affecting the distribution and structure composition of macrobenthic fauna in brackish water of coastal habitats (Mannino and Montagna [32], Ogunwenmo and Osuala [33], Nanami et al. [34]). Although Middle Songkhla Lake is not connected to the sea directly, this zone receives the effect of salinity from the saltwater inflow through the Lower Lake which is open to the Gulf of Thailand. Salinity is often regarded as a primary descriptor in estuarine ecosystems (Gaston [35], Lamptey and Armah [36]).

A sedimentary habitat contains information mirroring the functional biodiversity and activity patterns of macrobenthic fauna (Rosenberg et al. [37]). The main characteristics at the bottom of Middle Songkhla Lake are clay and silt (Angsupanich et al. [22]) except for station six, which is mainly sand (84.6%). We found that sand/clay excess was positively associated with the densities of Terebellidae, Aoridae, Pseudotanaidae, and Anthuridae, while a negative association was found with Pilargiidae. A typical genus *Sigambra* within Pilargiidae (Angsupanich et al. [22]) was found to be negatively related with sand-clay excess, a finding supported by a study in the southeastern Gulf of California reporting that the genus *Sigambra* was dominant in the areas of sand percentage of 1% or mud of 60-70% (Méndez [38]).

In addition, the genus *Marginella* within Marginellidae was also listed as being present in Middle Songkhla Lake (Angsupanich et al. [22]) thus showing a positive association with sand. This finding agrees with a study of invertebrate species identified in Fresh Creek, Bahamas where *Marginella* was listed as most commonly having the habitat type of sandflat (Layman and Silliman [39]).

Ten families (Nereididae, Stenothyridae, Nephtyidae, Isaeidae, Marginellidae, Alpheidae, Oedicerotidae, Apseudidae, Skeneopsidae, and Tellinidae) showed no evidence of association with any of the environmental variables. Although, Alpheidae was found to have the lowest density among the families included in our study, it is commonly found in the stomach contents of the dominant bottom feeding fish (*Osteogeneiosus militaris* and *Arius maculatus*) in Middle

Songkhla Lake. Angsupanich et al. [39] implied that these catfish species feed opportunistically on a variety of prey in their environment coupled with preferential feeding. So the low occurrence of Alpheidae may have been due to its swift movement and consequent catching difficulty.

Nereididae is one of the most important polychaete due to its diversity and abundance, found not only in marine environments (Gonzalez-Escalante and Salazar-Vallejo [40]) but also in brackish water such as occurs in Middle Songkhla Lake. Fourteen species of Nereididae were reported in a former study (Angsupanich et al. [22]) and it seems that Nereididae is widespread in Middle Songkhla Lake where it was the highest species richness. No evidence of Nereididae variation with salinity was found, possibly due to species diversity within this family. Some species, such as *Ceratonereis hircinicola*, were widely spread in the high salinity areas (Angsupanich and Kuwabara [41]), whereas *Namalycastis indica* has been found to inhabit fresh to slightly brackish water in cisterns, pools and lagoons (Glasby [42]).

Songkhla Lake nowadays suffers from the use of coastal land and water resources for uncontrolled shrimp farming, the destruction of both mangrove areas and peat swamp forest, construction of intake and outfall structures, and the construction of a deep sea port (Chufamane et al. [43]). The analytic methods we have used are designed to gain a better understanding of the environmental factors associated with macrobenthic fauna and can be used as an additional or alternative method for analysis of the relationships between environmental variables and the abundance of benthic organisms. This knowledge is useful for the natural resource management of estuarine environments.

ACKNOWLEDGEMENTS

We thank Professor Don McNeil for his advice on the statistical analysis. The data were collected using funds provided by the Biodiversity Research and Training Program (BRT), under the National Science and Technology Development Agency (NSTDA), together with the Thailand Research Fund (TRF) with the grant number 142016.

REFERENCES

- [1] Weisberg, S.B., Dauer, D.M., Schaffner, L.C., and Frithsen, J.B., An estuarine benthic index of biotic integrity (B-BI) for Chesapeake Bay, *Estuaries*, 1997; **20**: 149-158.
- [2] Borja, A., Franco, J., and Pérez, V., A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments, *Mar. Pollut. Bull.*, 2000; **40**: 1100-1114.
- [3] Ranasinghe, J.A., Thompson, B., Smith, R.W., Lowe, S., and Schiff, K.C., Evaluation of benthic assessment methodology in Southern California Bays and San Francisco Bay, Technical Report No. 432, Southern California Coastal Water Research Project, 2004.
- [4] Gray, J.S., Aschan, M., Carr, M.R., Clarke, K.R., Rosenberg, R., and Warwick, R.M., Analysis of community attributes of the benthic macrofauna of Frierfjord/ Langesundfjord and in a mesocosm experiment, *Mar. Ecol. Prog. Ser.*, 1988; **46**: 151-165.
- [5] Warwick, R.M., A new method for detecting pollution effects on marine macrobenthic communities, *Mar. Biol.*, **92**: 557-562.
- [6] Clarke, K.R., and Warwick, R.M., *Change in Marine Communities: an Approach to Statistical Analysis and Interpretation*, Natural Environment Research Council, Plymouth, UK, 1994.
- [7] Clarke, K.R., Somerfield, P.J., and Chapman, M.G., On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages, *J. Exp. Mar. Biol. Ecol.*, 2006; **330**: 55-80.
- [8] Ter Braak, C.J.F., Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis, *Ecology*, 1986; **67**: 1167-1179.

- [9] von Wehrden, H., Hanspach, J., Bruelheide, H., and Wesche, K., Pluralism and diversity-trends in the use and application of ordination methods 1990-2007, *J. Veg. Sci.*, 2009; **20**: 695-705.
- [10] Rakocinski, C.F., Brown, S.S., Gaston, G.R., Heard, R.W., Walker, W.W., and Summers, J.K., Macrobenthic responses to natural and contaminant-related gradients in Northern Gulf of Mexico estuaries, *Ecol. Appl.*, 1997; **7**: 1278-1298.
- [11] Hawkins, C.P., Norris, R.H., Hogue, J.N., and Feminella, J.W., Development and evaluation of predictive models for measuring the biological integrity of streams, *Ecol. Appl.*, 2000; **10**: 1456-1477.
- [12] Joy, M.K., and Death, R.G., Development and application of a predictive model of riverine fish community assemblages in the Taranaki region of the North Island, New Zealand, *N.Z. J. Mar. Freshwat. Res.*, 2000; **34**: 241-252.
- [13] Guerra-García, J.M., González-Villa, F.J., and García-Gómez, J.C., Aliphatic hydrocarbon pollution and macrobenthic assemblages in Ceuta harbour: a multivariate approach, *Mar. Ecol. Prog. Ser.*, 2003; **263**: 127-138.
- [14] Hajisamae, S., and Chou, L.M., Do shallow water habitats of an impacted coastal strait serve as nursery grounds for fish?, *Estuar. Coast. Shelf Sci.*, 2003; **56**: 281-290.
- [15] Morrisey, D.J., Turner, S.J., Mills, G.N., Williamson, R.B., and Wise, B.E., Factors affecting the distribution of benthic macrofauna in estuaries contaminated by urban runoff, *Mar. Environ. Res.*, 2003; **55**: 113-136.
- [16] Ysebaert, T., Herman, P.M.J., Meire, P., Craeymeersch, J., Verbeek, H., and Heip, C.H.R., Large-scale spatial patterns in estuaries: estuarine macrobenthic communities in the Schelde estuary, NW Europe, *Estuar. Coast. Shelf Sci.*, 2003; **57**: 335-355.
- [17] Quintino, V., Elliott, M., and Rodrigues, A.M., The derivation, performance and role of univariate and multivariate indicators of benthic change: Case studies at differing spatial scales, *J. Exp. Mar. Biol. Ecol.*, 2006; **330**: 368-382.
- [18] Anderson, M.J., Animal-sediment relationships re-visited: Characterising species distributions along an environmental gradient using canonical analysis and quantile regression splines, *J. Exp. Mar. Biol. Ecol.*, 2008; **366**: 16-27.
- [19] Glockzin, M., and Zettler, M.L., Spatial macrozoobenthic distribution patterns in relation to major environmental factors- A case study from the Pomeranian Bay (southern Baltic Sea), *J. Sea. Res.*, 2008; **59**: 144-161.
- [20] Liang, S., Shieh, B., and Fu, Y., A structural equation model for physiochemical variables of water, benthic invertebrates, and feeding activity of waterbirds in the Sitsao wetlands of Southern Taiwan, *Zool. Stud.*, 2002; **41**: 441-451.
- [21] Warton, D.I., and Hudson, H.M., A MANOVA statistic is just as powerful as distance-based statistics, for multivariate abundances, *Ecology*, 2004; **85**: 858-874.
- [22] Angsupanich, S., Siripech, A., and Charoenpornthip, M., Macrobenthic fauna community in the Middle Songkhla Lake, Southern Thailand, *Songklanakarin J. Sci. Technol.*, 2005(a); **27**: 365-390.
- [23] Ter Braak, C.J.F., and Šmilauer, P., CANOCO Reference manual and CanoDraw for Windows User's guide: Software for Canonical Community Ordination (version 4.5), Microcomputer Power, Ithaca, NY, USA, 2002.
- [24] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org>, 2009.
- [25] Browne, M.W., An overview analytic rotation in exploratory factor analysis, *Multivariate. Behav. Res.*, 2001; **36**: 111-150.
- [26] Abdi, H., Factor rotations in factor analyses. In: Lewis-Beck, M., A. Bryman, T. Futing (Eds), *Encyclopedia of Social Sciences Research Methods*, Thousand Oaks, CA, Sage, 2003: 1-8.
- [27] Mardia, K.V., Kent, J.T., and Bibby, J.M., *Multivariate Analysis*, Academic Press, USA, 1979.
- [28] Venables, W.N., and Ripley, B.D., *Modern Applied Statistics with S*, 4th Edn., Springer Science + Business Media, New York, USA, 2002.

- [29] Olson, C.L., On choosing a test statistic in multivariate analysis of variance, *Psychol. Bull.*, 1976; **83**: 579-586.
- [30] Johnson, R.A., and Wichern, D.W., *Applied Multivariate Statistical Analysis*, 4th Edn., Prentice-Hall, USA, 1998.
- [31] Day, R.L., and Blake, J.A., Reproduction and larval development of *Polydora Giardi* Mesnil (Polychaeta: Spionidae), *Biol. Bull.*, 1979; **156**: 20-30.
- [32] Mannino, A., and Montagna, P.A., Small-scale spatial variation of macrobenthic community structure, *Estuaries*, 1997; **20**: 159-173.
- [33] Ogunwenmo, C.A., and Osuala, I.A., Physico-chemical parameters and macrobenthos of an estuarine creek and an artificial pond in Lagos, South-Western Nigeria, *Acta SATECH*, 2004; **1**: 128-132.
- [34] Nanami, A., Saito, H., Akita, T., Motomatsu, K., and Kuwahara, H., Spatial distribution and assemblage structure of macrobenthic invertebrates in a brackish lake in relation to environmental variables, *Estuar. Coast. Shelf Sci.*, 2005; **63**: 167-176.
- [35] Gaston, G.R., Lee, D.A.L., and Nasci, J.C., Estuarine macrobenthos in Calcasieu Lake, Louisiana: Community and trophic structure, *Estuaries and Coasts*, 1988; **11**: 192-200.
- [36] Lamptey, E., and Armah, A.L., Factors affecting macrobenthic fauna in a tropical hypersaline coastal lagoon in Ghana, West Africa, *Estuaries and Coasts*, 2008; **31**: 1006-1019.
- [37] Rosenberg, R., Magnusson, M., and Nilsson, H.C., Temporal and spatial changes in marine benthic habitats in relation to the EU Water Framework Directive: The use of sediment profile imagery, *Mar. Pollut. Bull.*, 2009 ; **58**: 565-572.
- [38] Méndez, N., Relationships between deep-water polychaete fauna and environmental factors in the southeastern Gulf of California, Mexico, *Sci. Mar. (Barc.)*, 2007; **71**: 605-622.
- [39] Angsupanich, S., Somsak, S., and Phrommoon, J., Stomach contents of the catfishes *Osteogobius militaris* (Linnaeus, 1758) and *Arius maculatus* (Thunberg, 1792) in the Songkhla Lake, *Songklanakarini J. Sci. Technol.*, 2005(b); **27**: 391-402.
- [39] Layman, C.A., and Silliman, B.R., Preliminary survey and diet analysis of juvenile fishes of an estuarine creek on Andros Island, Bahamas, *Bull. Mar. Sci.*, 2002; **70**: 199-210.
- [40] Gonzalez-Escalante, L.E., and Salazar-Vallejo, S.I., A new estuarine species, *Nereis garwoodi* (Polychaeta: Nereididae), from Bahia Chetumal, Mexican Caribbean coast, *Rev. Biol. Trop.*, 2003; **51**: 155-164.
- [41] Angsupanich, S., and Kuwabara, R., Macrobenthic fauna in Thale Sap Songkla, a brackish lake in southern Thailand, *Lakes Reservoirs: Res. Manage.*, 1995; **1**: 115-125.
- [42] Glasby, C.H., *The Namanereidinae (Polychaeta: Nereididae), Part 1, Taxonomy and phylogeny*, Records of the Australian Museum, 1999: 129 pp.
- [43] Chufamane, P., Boromthanasat, S., and Lønholdt, J., A case experience on integrated environment and water management towards people's livelihoods, Case story 1: Partnership policy in Songkhla Lake Basin, Workshop report. Songkhla workshop on linking management of catchments and coastal ecosystems, Royal Thai Ministry of Natural Resources and Environment, 2003: 15-32.

Appendix 3

Publication 3: Confidence intervals for adjusted proportions using logistic regression

Prince of Songkla University
Pattani Campus

Confidence Intervals for Adjusted Proportions Using Logistic Regression

Noodchanath Kongchouy (Corresponding author)

Department of Mathematics, Faculty of Science and Technology,
Prince of Songkla University, Hat Yai, Songkhla, 90112, Thailand

Tel: 66-84-063-1251 E-mail: nootchanath.k@psu.ac.th

Uraiwan Sampantarak

Pattani Inland Fisheries Research and Development Center,

Department of Fisheries, Ministry of Agricultural and Cooperatives, Pattani 94160, Thailand

Tel: 66-84-653-9391 E-mail: uraiwans@fisheries.go.th; uraiwan111@hotmail.com

This study was partially funded by the Graduate School, Prince of Songkla University.

Abstract

This paper presents confidence intervals for adjusted proportions using logistic regression with weighted sum contrasts. The methods are applied to data from two studies, (1) imposex percentages among female gastropods at different locations in the Gulf of Thailand adjusted for different species, and (2) complication-based neonatal morbidity risk for births at a major hospital adjusted for demographic factors.

Keywords: Confidence interval, Proportion, Logistic regression, Sum

1. Introduction

Odds ratios are conventionally used for assessing associations between binary outcomes and categorical risk factors. They are often preferred in scientific studies because they give valid confidence intervals for these associations for case-control studies as well as for cohort studies and cross-sectional studies (Fernandez et al., 1999; Lim & Tongkumchum, 2009; Peters et al., 2000). A further advantage is that methods such as Mantel-Haenszel adjustment and logistic regression are available for adjusting odds ratios for confounding bias arising from covariates associated with both the binary outcome and the risk factor of interest. These issues are discussed in detail in the biostatistical literature (see, for example, McNeil, 1996; Woodward, 1999).

For cohort and cross-sectional studies where the proportions or percentages of an adverse outcome are of primary interest, it is also important to give confidence intervals for comparing these proportions. If there are no covariates of interest, these confidence intervals can be computed, either directly from the observed proportions with some adjustment such as an arcsine transformation to ensure that the confidence intervals are between 0 and 1 (see, for example, Armitage & Berry, 1994) using a logistic regression model. However, the situation is more complicated when adjustments for covariates are required, and methods for constructing such confidence intervals are not routinely provided in statistical packages.

In this paper we describe a general method for computing confidence intervals for comparing several proportions after adjusting for categorical covariates. The method is illustrated using two recently published applications of scientific interest: imposex among female gastropods in the Gulf of Thailand (Swennen, Sampantarak, & Rattanadakul, 2009), and complication-based neonatal morbidity risk among babies at a major hospital in southern Thailand (Rachatapantanakorn & Tongkumchum, 2009).

2. Methods

2.1 Logistic Regression

Logistic regression (Hosmer & Lemeshow, 2000; Kleinbaum & Klein, 2002) is a statistical method widely used to model the association between a binary outcome probability - the probability of a

specific outcome - and a set of fixed determinants. When the determinants are categorical factors, these factors can be structured as a multi-way contingency table of counts and the data for analysis comprise the proportions of adverse outcomes in the cells of this table. If, for example, the outcome variable Y takes values 0 and 1 (adverse outcome) and there are two factors with levels indexed by i and j , respectively, the model takes the form

$$\text{Prob}[Y_{ij} = 1] = 1 / (1 + e^{-(\alpha_i + \beta_j)}). \quad (1)$$

Note that if r and c are the numbers of factor levels specified in the model by the sets of parameters $\{\alpha\}$ and $\{\beta\}$, the number of independent parameters is $r + c - 1$, so it is necessary to put a constraint on these parameters when fitting the model. This constraint is conventionally achieved by replacing $\alpha_i + \beta_j$ in the model by $\mu + \alpha_i + \beta_j$ where $\alpha_1 = 0$ and $\beta_1 = 0$, in which case the model is written as

$$\text{Prob}[Y_{ij} = 1] = 1 / (1 + e^{-(\mu + \alpha_i + \beta_j)}). \quad (2)$$

This model provides estimates of odds ratios for comparing the outcome probabilities with respect to specified levels of each factor. Thus $\exp(\alpha_i)$ is the ratio of the odds of an adverse outcome for level i to that for level 1 for the first factor, whereas $\exp(\beta_j)$ is the ratio of the odds of an adverse outcome for level j to that for level 1 for the second factor. Thus each odds ratio in this model uses the first level of each factor as a baseline. To obtain odds ratios with respect to another baseline level, the data would need to be recoded so that the new baseline is constrained to take the value 0.

Under appropriate conditions on the pattern of zeros in the data (see, for example, Section 7.2 of Venables & Ripley, 2002), this logistic regression model is fitted using maximum likelihood and the results include estimates of the parameters and their standard errors, from which confidence intervals can be plotted.

2.2 Contrasts

When the constraints $\alpha_1 = 0$ and $\beta_1 = 0$ are used in the logistic regression model (2) the confidence intervals apply to the *differences* between each of the sets of parameters and the first parameter specified in each factor. These differences are known as *treatment contrasts*. In practice, it is often preferable not to single out a specific level of a factor as a basis for comparison, but rather to treat all factor levels in the same way. For linear models of the form

$$Y_{ij} = \alpha_i + \beta_j + \varepsilon_{ij}, \quad (3)$$

standard errors of these differences are obtained by using the standard sum contrasts available in commonly used software packages such as R program (R Development Core Team, 2007). However, as pointed out by Venables and Ripley (2002), these contrasts are not valid for unbalanced designs, which include logistic regression models. Thus it is necessary to construct specific contrasts for logistic regression, and this can be accomplished by using weighted sum contrasts rather than treatment contrasts (Tongkumchum & McNeil, 2009). These weighted sum contrasts provide standard errors for the differences between each factor level and their overall mean.

2.3 Adjustment for Covariates

The method is analogous to that commonly used by linear regression analysis for adjusting outcomes to reduce the effects of covariate factors, such as seasonal adjustment of unemployment rates. In this case, if $\{\beta\}$ represents the primary factor of interest and $\{\alpha\}$ the covariate factor, and ε_{ij} are independent errors with mean zero and common standard deviation, the model is given by Equation (3). The factors $\{\alpha\}$ and $\{\beta\}$ in this model are estimated as the row and column means of the data matrix y_{ij} , with a suitable constraint to ensure identifiability. To adjust for the covariate, the adjusted mean for level j of the primary factor is obtained by first removing the effect of the covariate from each observation by replacing y_{ij} by $y_{ij} - \hat{\alpha}_i$ and then adding a constant to ensure that the mean of the corrected observations remains the same as the mean of the original observations. As a result, the adjusted mean is $\bar{y}_j = \hat{\beta}_j + d$, where d is a constant chosen to ensure that the overall mean before and after the adjustment remains the same. It follows that $d = \bar{y} - \bar{\beta}$, where $\bar{\beta}$ is the mean of the estimated β parameters.

Similarly, the formula for adjusting the proportion for level j of a primary factor is

$$p_j^* = 1 / (1 + e^{-(\beta_j + d)}). \quad (4)$$

Note that this result follows from the fact that the estimate given by logistic regression for the adjusted odds ratio for level j compared to level 1 is $\exp(\hat{\beta}_j)$ and this must equate to $\left(\frac{p_j^*}{1-p_j^*}\right) \bigg/ \left(\frac{p_1^*}{1-p_1^*}\right)$.

The constant d may be chosen to ensure that the overall proportion (or the total number N) of adverse outcomes before and after the adjustment is the same, that is

$$\sum_{i=1}^r \sum_{j=1}^c n_{ij} p_j^* = \sum_{i=1}^r \sum_{j=1}^c n_{ij} p_{ij} = N. \tag{5}$$

Substituting Equation (4) into Equation (5), it follows that d must satisfy the equation

$$\sum_{j=1}^c \left(\sum_{i=1}^r n_{ij} \right) \bigg/ \left(1 + e^{-(\beta_j + d)} \right) = N. \tag{6}$$

Equation (6) is non-linear and cannot be solved explicitly to give an expression for the constant d . However, it can be solved using the Newton-Raphson iterative procedure with Marquardt damping to ensure convergence. Note that this method extends straightforwardly to any number of covariate factors.

3. Applications

3.1 *Imposex among gastropods in the Gulf of Thailand*

Female gastropods were collected from 56 sampling sites grouped into 13 areas around the Gulf of Thailand in 2006 (Swennen et al., 2009) and tested for imposex. Since different species have different sensitivities to be imposex and this variation could bias the estimation of imposex prevalence due to the fact that different species are found in different locations, it was necessary to take both factors (16 species groups and 13 areas) into account. The logistic regression model (Equation 2) was fitted to the data. Figure 1 shows plots of 95% confidence intervals for the proportions.

The overall imposex percentage was 25.2, indicated by the vertical lines in Figure 1. Although there were 208 (13 areas \times 16 species groups) cells in the data table, 124 of these cells contained no data because relatively few species groups were found in each area. Thus the number of degrees of freedom after fitting the two-factor model was reduced from 71 (84 - 13) to 56 (84 - 13 - 15).

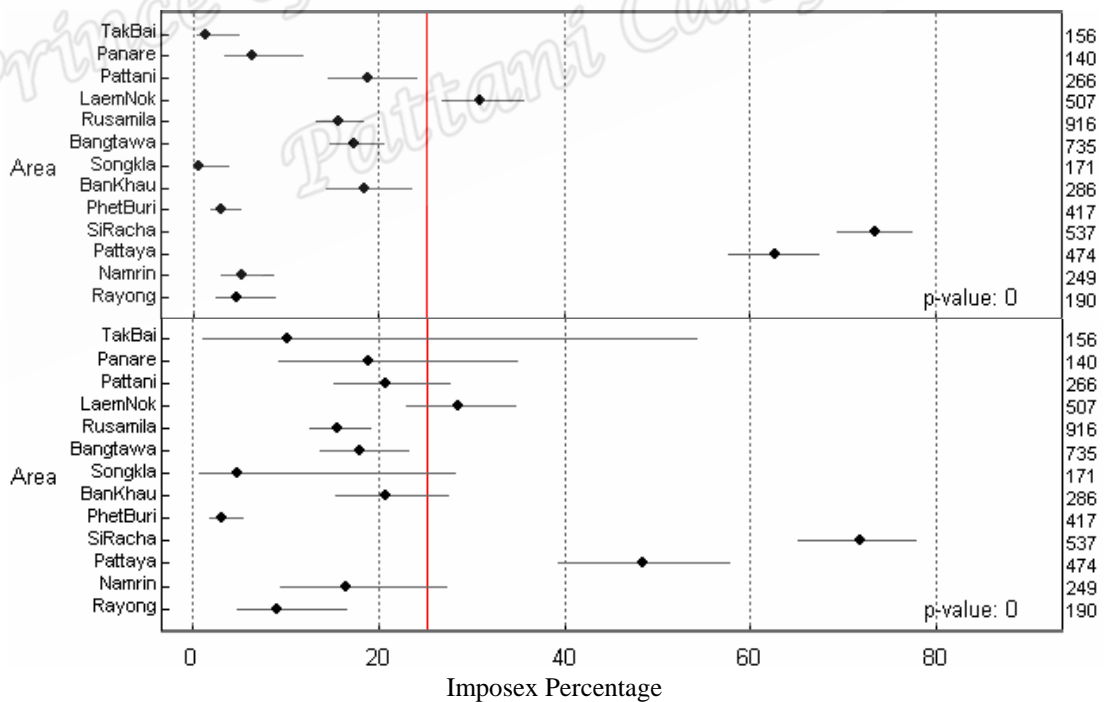


Figure 1. The 95% confidence intervals for percentages of female gastropods with imposex disease at various locations around the Gulf of Thailand; the upper and lower panels show crude and species-adjusted percentages, and the sample sizes are given on the right.

Figure 1 indicates that the crude percentages overestimated the true values in the Pattaya area and underestimated them in the Tak Bai, Pattani and Namrin areas. Due to the fact that 15 additional parameters were needed to take the species factor into account, the adjusted percentages have wider confidence intervals than the crude percentages. This effect is particularly notable for the Tak Bai area because the imposex was only found to occur among one of the three species groups observed there.

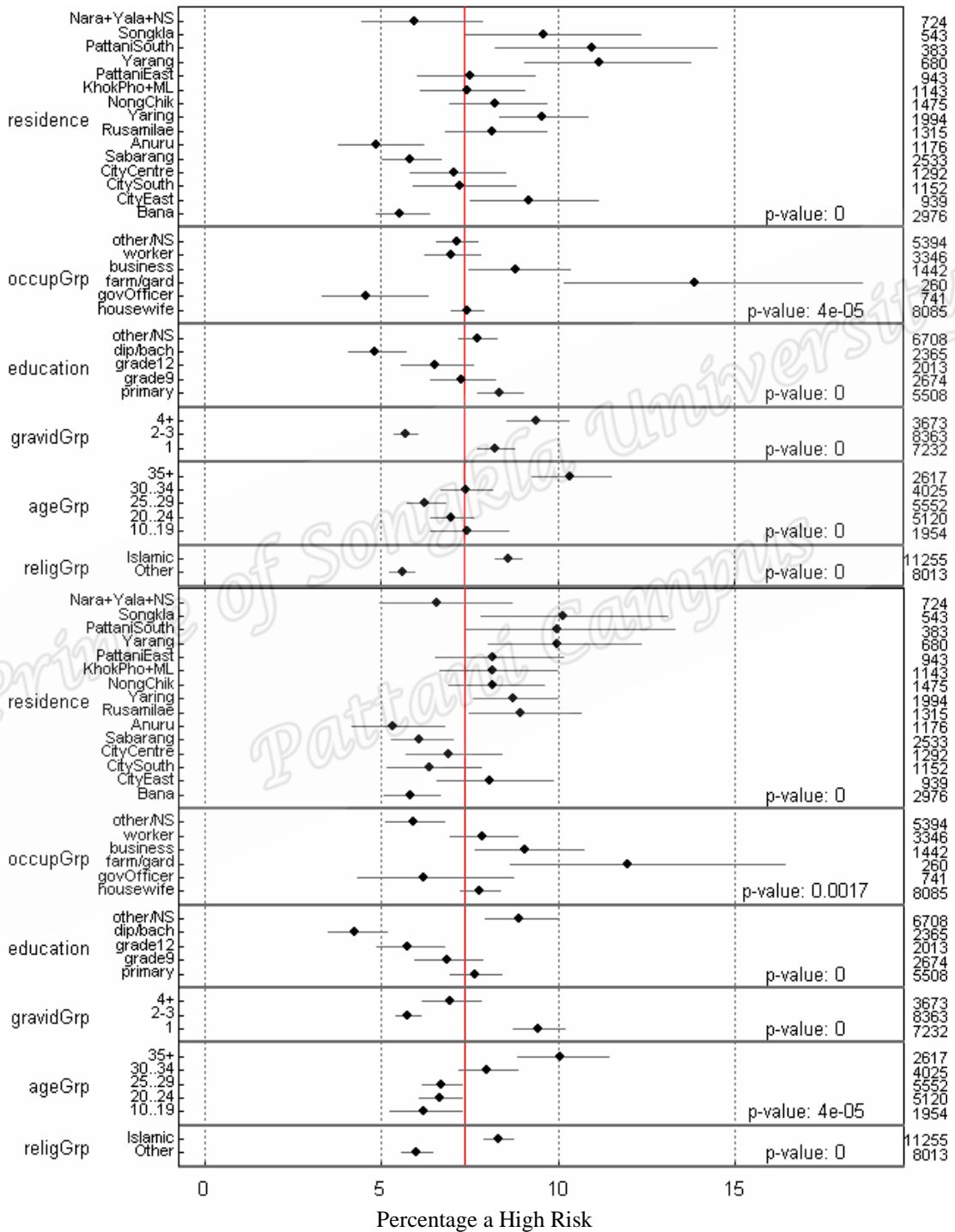


Figure 2. The 95% confidence intervals for complication-based risks among singleton deliveries to mothers with no previous Caesarean-section birth at Pattani Hospital, with respect to each of six demographic factors before (upper panel) and after (lower panel) adjustment for other factors.

3.2 Complication-based neonatal morbidity risk

Based on complications recorded for 19,268 singleton deliveries to mothers with no previous Caesarean-section birth at Pattani Hospital in Southern Thailand over a nine-year period from 1997 to 2005, Rachatapantanakorn and Tongkumchum (2009) classified babies as high or low risk, and used logistic regression to assess the effects of six demographic risk factors.

Figure 2 shows 95% confidence intervals of the crude and adjusted percentages for each determinant before and after adjusting for the other determinants. The main differences between the crude and adjusted risks occurred among three demographic factors: the number of pregnancies (gravid group), age-group and religion.

Mothers in gravid group 4 or more pregnancies had highest risk percentage before adjustment, but after adjusting for the other determinants the risk was highest for primigravid mothers. For the age-group effect, the crude percentage was lowest for mothers aged 25-29 years, whereas the adjusted percentage was lowest for mothers aged less than 20 years. It is also noteworthy that the quite large difference in risks between Muslim and non-Muslim mothers was reduced after adjusting for the other factors.

4. Discussion

In this paper we have described a simple method for adjusting proportions for categorical covariates based on a fitted logistic regression model that provides asymptotically valid confidence intervals for comparing proportions over different levels of a categorical risk factor. While this method is not entirely new (see, for example, related earlier work by Berthold et al., 2007; Graubard & Korn, 1999; Lane & Nelder, 1982), it is not widely used in scientific studies, particularly when comparing more than two proportions. A further advantage of the method is that by using appropriately weighted sum contrasts each proportion can be compared with the overall mean rather than with a specified reference group.

5. Acknowledgement

We are grateful to Professor Don McNeil for his helpful guidance.

References

- Armitage, P., & Berry, G. (1994). *Statistical methods in medical research*. (3rd ed.). UK: Blackwell-science, Oxford.
- Berthold, S. M., Wong, E. C., Schell, T. L., Marshall, G. N., Elliott, M. N., Takeuchi, D., & Hambarsoomians, K. (2007). U.S. Cambodian refugees' use of complementary and alternative medicine for mental health problems. *Psychiatric Services*, 58, 1212-1218.
- Fernandez, E., Schiaffino, A., Rajmil, L., Badia, X., & Segura, A. (1999). Gender inequalities in health and health care services use in Catalonia (Spain). *Journal of Epidemiology and Community Health*, 53(4), 218-222.
- Graubard, B. I., & Korn, E. L. (1999). Predictive margins with survey data, *Biometrics*, 55, 652-659.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. (2nd ed.). John Wiley and Sons.
- Kleinbaum, D. G., & Klein, M. (2002). *Logistic Regression: a Self-Learning Text*. (2nd ed.). New York: Springer-Verlag.
- Lane, P. W., & Nelder, J. A. (1982). Analysis of covariance and standardization as instances of prediction, *Biometrics*, 38(3), 613-621.
- Lim, A., & Tongkumchum, P. (2009). Methods for analyzing hospital length of stay with application to inpatients dying in Southern Thailand. *Global Journal of Health Science*, 1(1), 27-38.
- McNeil, D. (1996). *Epidemiological research methods*. Chichester: John Wiley & Sons.
- Peters, A., Liu, E., Verrier, R. L., Schwartz, J., Gold, D. R., Mittleman, M., Baliff, J., Oh, A. J., Allen, G., Monahan, K., & Dockery, D. W. (2000). Air pollution and incidence of cardiac arrhythmia. *Epidemiology*, 11, 11-17.
- R Development Core Team. (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.

Rachatapantanakorn, O., & Tongkumchum, P. (2009). Demographic determinants for caesarean delivery in Pattani hospital. *Southeast Asian Journal of Tropical Medicine and Public Health*, 40, 602-611.

Swennen, C., Sampantarak, U., & Rattanadakul, N. (2009). TBT-pollution in the Gulf of Thailand: a re-inspection of imposex incidence after 10 years. *Marine Pollution Bulletin*, 58, 526-532.

Tongkumchum, P., & McNeil, D. (2009). Confidence intervals using contrasts for regression model. *Songklanakarin Journal of Science and Technology*, 31, 151-156.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. (4th ed.). New York: Springer Science + Business Media, (Section 6.2).

Woodward, M. (1999). *Epidemiology: study design and data analysis*. USA: Chapman & Hall.

Prince of Songkla University
Pattani Campus

Appendix 4: Names of the study areas indicated by a letter and the name of a town or village on the adjacent coast, the positional coordinates, average depth of the sampling sites, the names of the species, number of males, number of females, and number of imposex in 2006

Study area	Site	Position	Depth (m)	Species	Males	Females	Imposex			
A: Rayong	1	12° 35' 00" N 101° 29' 00" E	8-20	<i>Murex altispira</i>	8	31	0			
				<i>Murex trapa</i>	0	4	0			
				<i>Thais lacera</i>	2	3	3			
				<i>Babylonia areolata</i>	11	20	0			
				<i>Pugilina cochlidium</i>	75	52	2			
				<i>Hemifusus ternatanus</i>	3	2	0			
	2	12° 37' 47" N 101° 26' 46" E	0	<i>Semiricinula muricoides</i>	37	68	3			
				<i>Thais rufotincta</i>	12	10	1			
B: Namrin	3	12° 40' 39" N 101° 06' 47" E	0	<i>Semiricinula muricoides</i>	3	19	3			
				<i>Thais clavigera</i>	9	15	9			
				<i>Morula musiva</i>	84	87	0			
	4	12° 40' 00" N 101° 02' 48" E	0	<i>Semiricinula muricoides</i>	15	33	0			
				<i>Thais clavigera</i>	30	21	0			
				<i>Morula musiva</i>	79	74	1			
C: Pattaya	5	12° 58' 14" N 100° 53' 30" E	0	<i>Lataxiena blosvillei</i>	203	242	214			
				6	12° 59' 45" N 100° 55' 22" E	0	<i>Semiricinula muricoides</i>	45	155	68
							7	13° 00' 18" N 100° 55' 24" E	0	<i>Pugilina cochlidium</i>
D: Si Racha	8	13° 07' 39" N 100° 53' 54" E	0	<i>Semiricinula muricoides</i>	32	126	109			
				9	13° 09' 30" N 100° 54' 00" E	6	<i>Murex trapa</i>	2	3	3
							<i>Pugilina cochlidium</i>	9	6	4
	<i>Hemifusus ternatanus</i>	31	59				16			
	10	13° 09' 00" N 100° 54' 27" E	0	<i>Semiricinula muricoides</i>	100	120	82			
				11	13° 12' 16" N 100° 56' 00" E	0	<i>Lataxiena blosvillei</i>	152	124	111
<i>Pugilina cochlidium</i>	0	1	1							
12	13° 10' 00" N 100° 50' 00" E	4	<i>Murex trapa</i>	15	7	5				
			<i>Hemifusus ternatanus</i>	1	3	0				

Appendix 4: (Cont.)

Study area	Site	Position	Depth (m)	Species	Males	Females	Imposex
	13	13° 10' 00" N 100° 49' 00" E	8	<i>Murex trapa</i>	35	37	27
	14	13° 10' 00" N 100° 47' 00" E	12-20	<i>Murex trapa</i>	51	50	36
				<i>Nassarius livescens</i>	1	0	0
				<i>Hemifusus ternatanus</i>	1	1	1
				<i>Turricula javana</i>	1	0	0
E: Phet Buri	15	13° 06' 00" N 100° 04' 00" E	5	<i>Nassarius stolatus</i>	86	78	0
	16	13° 06' 00" N 100° 07' 00" E	2-5	<i>Thais lacera</i>	52	62	1
				<i>Babylonia areolata</i>	1	0	0
				<i>Pugilina cochlidium</i>	4	4	0
				<i>Hemifusus ternatanus</i>	1	3	0
				<i>Turricula javana</i>	0	1	0
	17	13° 05' 00" N 100° 12' 00" E	6	<i>Thais lacera</i>	18	21	0
				<i>Pugilina cochlidium</i>	12	11	0
				<i>Hemifusus ternatanus</i>	1	0	0
	18	13° 03' 00" N 100° 16' 00" E	10-30	<i>Murex altispira</i>	0	2	0
				<i>Murex trapa</i>	146	232	12
				<i>Pugilina cochlidium</i>	1	0	0
				<i>Hemifusus ternatanus</i>	0	3	0
F: Ban Khau	19	12° 53' 00" N 100° 03' 00" E	6	<i>Murex trapa</i>	16	47	22
				<i>Thais lacera</i>	5	13	2
				<i>Pugilina cochlidium</i>	7	12	0
				<i>Hemifusus ternatanus</i>	44	61	0
	20	12° 53' 00" N 100° 06' 00" E	9-10	<i>Murex trapa</i>	47	91	17
				<i>Thais lacera</i>	4	6	0
				<i>Pugilina cochlidium</i>	20	22	0
	21	12° 53' 00" N 100° 10' 00" E	10-14	<i>Murex trapa</i>	2	13	7
				<i>Thais lacera</i>	19	15	5
	22	12° 10' 00" N 100° 10' 00" E	8-10	<i>Murex trapa</i>	2	5	0
				<i>Thais lacera</i>	2	0	0
				<i>Turricula javana</i>	0	1	0

Appendix 4: (Cont.)

Study area	Site	Position	Depth (m)	Species	Males	Females	Imposex
G: Songkla	23	07° 15' 00" N 100° 46' 00" E		<i>Murex altispira</i>	33	150	0
				<i>Rapana rapiformis</i>	30	19	1
				<i>Hemifusus ternatanus</i>	0	2	0
H: Bang Tawa	24	6° 52' 00" N 101° 04' 20" E	2	<i>Murex altispira</i>	0	1	0
				<i>Murex occa</i>	6	7	0
				<i>Murex trapa</i>	15	18	1
				<i>Thais lacera</i>	3	3	0
				<i>Nassaria pusilla</i>	12	44	1
				<i>Nassarius siquijorensis</i>	25	50	3
				<i>Nassarius jacksonianus</i>	0	3	0
				<i>Hemifusus ternatanus</i>	0	2	1
	<i>Turricula javana</i>	32	68	3			
	25	6° 52' 00" N 101° 07' 25" E	2	<i>Murex occa</i>	1	1	0
				<i>Thais lacera</i>	1	3	1
				<i>Nassarius siquijorensis</i>	4	26	13
				<i>Hemifusus ternatanus</i>	0	5	2
<i>Turricula javana</i>				0	6	2	
26	6° 51' 37" N 101° 09' 30" E	0	<i>Thais bitubercularis</i>	208	208	30	
27	6° 52' 00" N 101° 09' 30" E	2	<i>Murex occa</i>	2	2	0	
			<i>Thais lacera</i>	3	3	0	
			<i>Babylonia areolata</i>	0	4	0	
			<i>Nassaria pusilla</i>	2	9	0	
			<i>Nassarius siquijorensis</i>	14	103	40	
			<i>Nassarius jacksonianus</i>	1	6	0	
			<i>Pugilina cochlidium</i>	5	9	0	
			<i>Hemifusus ternatanus</i>	1	11	4	
<i>Turricula javana</i>	22	101	26				
28	6° 58' 30" N 101° 10' 00" E	8-10	<i>Murex altispira</i>	2	3	0	
			<i>Murex trapa</i>	2	0	0	
			<i>Rapana rapiformis</i>	1	0	0	
			<i>Hemifusus ternatanus</i>	1	0	0	
29	6° 52' 49" N 101° 10' 00" E	3-4	<i>Murex occa</i>	1	2	0	
			<i>Murex trapa</i>	0	1	1	

Appendix 4: (Cont.)

Study area	Site	Position	Depth (m)	Species	Males	Females	Imposex
				<i>Pugilina cochlidium</i>	9	4	0
				<i>Hemifusus ternatanus</i>	35	32	0
I: Rusamilae	30	6° 53' 48" N 101° 13' 55" E	3	<i>Murex occa</i>	16	19	2
				<i>Thais lacera</i>	12	6	0
				<i>Nassarius siquijorensis</i>	2	1	0
				<i>Nassarius jacksonianus</i>	19	42	5
				<i>Pugilina cochlidium</i>	0	1	0
				<i>Hemifusus ternatanus</i>	12	8	0
				<i>Turricula javana</i>	19	41	0
	31	6° 53' 20" N 101° 13' 00" E	3	<i>Murex occa</i>	20	22	3
				<i>Nassarius jacksonianus</i>	36	102	12
				<i>Nassarius livescens</i>	13	8	1
				<i>Pugilina cochlidium</i>	5	5	0
				<i>Hemifusus ternatanus</i>	7	6	0
				<i>Turricula javana</i>	33	54	1
	32	6° 53' 30" N 101° 13' 30" E	2	<i>Murex trapa</i>	125	182	28
				<i>Thais lacera</i>	25	29	2
				<i>Hemifusus ternatanus</i>	2	18	9
				<i>Turricula javana</i>	6	33	7
	33	6° 53' 30" N 101° 14' 00" E	3-5	<i>Murex occa</i>	130	48	19
	34	6° 55' 00" N 101° 13' 00" E	4	<i>Murex occa</i>	41	64	23
				<i>Thais lacera</i>	0	1	0
				<i>Nassarius siquijorensis</i>	7	7	0
				<i>Nassarius stolatus</i>	68	151	14
				<i>Turricula javana</i>	11	68	17
J: Laem Nok	35	6° 52' 40" N 101° 15' 30" E	2	<i>Murex altispira</i>	0	1	0
				<i>Murex occa</i>	67	43	13
				<i>Murex trapa</i>	0	2	0
				<i>Thais bitubercularis</i>	1	3	0
				<i>Nassarius siquijorensis</i>	5	14	2
				<i>Nassarius jacksonianus</i>	9	12	1
				<i>Pugilina cochlidium</i>	1	2	0
				<i>Hemifusus ternatanus</i>	1	3	1

Appendix 4: (Cont.)

Study area	Site	Position	Depth (m)	Species	Males	Females	Imposex
				<i>Turricula javana</i>	5	14	5
	36	6° 53' 00" N 101° 18' 00" E	1.5	<i>Murex occa</i>	4	6	4
				<i>Thais lacera</i>	59	21	5
				<i>Pugilina cochlidium</i>	78	69	3
	37	6° 53' 30" N 101° 18' 30" E	2	<i>Murex occa</i>	14	16	8
				<i>Thais lacera</i>	17	11	1
				<i>Pugilina cochlidium</i>	90	50	1
	38	6° 54' 00" N 101° 19' 30" E	1-3	<i>Murex occa</i>	0	1	1
				<i>Thais lacera</i>	22	22	4
	39	6° 54' 00" N 101° 17' 00" E	1.5	<i>Murex occa</i>	72	90	46
				<i>Murex trapa</i>	0	2	0
				<i>Thais lacera</i>	5	28	19
				<i>Pugilina cochlidium</i>	72	48	10
	40	6° 53' 30" N 101° 17' 30" E	3-4	<i>Murex occa</i>	45	35	29
				<i>Thais lacera</i>	27	14	4
K: Pattani Bay Mouth	41	6° 54' 30" N 101° 15' 30" E	4	<i>Murex occa</i>	4	8	7
				<i>Thais lacera</i>	1	6	5
				<i>Pugilina cochlidium</i>	5	2	1
	42	6° 55' 00" N 101° 15' 00" E	2-3	<i>Murex occa</i>	31	33	15
				<i>Thais lacera</i>	2	36	15
				<i>Pugilina cochlidium</i>	51	34	2
	43	6° 55' 16" N 101° 14' 20" E	5	<i>Murex occa</i>	12	13	0
				<i>Nassarius livescens</i>	14	13	0
				<i>Pugilina cochlidium</i>	5	1	0
				<i>Hemifusus ternatanus</i>	2	5	0
				<i>Turricula javana</i>	4	12	2
	44	6° 55' 24" N 101° 14' 16" E	5	<i>Murex occa</i>	5	7	0
				<i>Thais lacera</i>	2	2	0
				<i>Nassarius jacksonianus</i>	2	1	0
				<i>Nassarius livescens</i>	3	6	1
				<i>Pugilina cochlidium</i>	3	3	0
				<i>Hemifusus ternatanus</i>	0	1	1
				<i>Turricula javana</i>	2	6	0

Appendix 4: (Cont.)

Study area	Site	Position	Depth (m)	Species	Males	Females	Imposex
	45	6° 56' 05" N 101° 14' 55" E	0	<i>Nassarius stolatus</i>	73	77	1
L: Panare	46	7° 01' 00" N 101° 14' 00" E	10	<i>Murex altispira</i>	0	1	0
				<i>Murex trapa</i>	0	4	1
				<i>Thais lacera</i>	0	1	0
				<i>Babylonia areolata</i>	0	3	0
	47	7° 02' 00" N 101° 16' 00" E	10-15	<i>Nassarius livescens</i>	2	2	0
	48	7° 08' 00" N 101° 15' 00" E	20	<i>Phos senticosus</i>	1	1	0
				<i>Nassarius siquijorensis</i>	1	1	0
	49	6° 54' 45" N 101° 23' 00" E	8-10	<i>Murex altispira</i>	1	0	0
				<i>Chicoreus banksii</i>	0	2	0
				<i>Rapana rapiformis</i>	0	1	0
	50	7° 03' 00" N 101° 32' 00" E	20-30	<i>Murex altispira</i>	5	30	1
				<i>Chicoreus banksii</i>	7	8	1
	51	6° 54' 50" N 101° 28' 00" E	10	<i>Murex altispira</i>	6	12	0
				<i>Rapana rapiformis</i>	1	1	0
	52	6° 57' 30" N 101° 30' 00" E	20-25	<i>Murex altispira</i>	5	29	0
				<i>Chicoreus banksii</i>	7	24	4
				<i>Rapana rapiformis</i>	21	11	1
	53	6° 57' 00" N 101° 32' 00" E	20	<i>Murex altispira</i>	6	4	1
				<i>Chicoreus banksii</i>	5	5	0
M: Tak Bai	54	6° 16' 50" N 102° 03' 30" E	4-6	<i>Babylonia areolata</i>	144	131	2
	55	6° 14' 50" N 102° 05' 10" E	0	<i>Thais bitubercularis</i>	4	3	0
	56	6° 20' 00" N 102° 09' 00" E	15-20	<i>Murex altispira</i>	7	22	0
					3713	5044	1270