Chapter 4

A multivariate multiple regression model for macrobenthic fauna density

This chapter presents a preliminary analysis and a model fitting macrobenthic fauna distribution in the Middle Songkhla Lake. Section 4.1 is the descriptions of the variables. Section 4.2 is data characteristics. Section 4.3 is the multivariate multiple regression (MMR) model fitted to the density of twenty-four families of macrobenthic fauna (see more details, Publication 2, in Appendix 2).

4.1 Description of the variables

The roles of variables are classified as determinants and outcome. These variables, their roles and data type as shown in Table 4.1. The outcome of interest is the densities of the twenty-four families of macrobenthic fauna from nine sampling stations at six bimonthly periods. These densities are of continuous data type. The predictor variables (determinants) consisted of the environmental factors and the unique variable derived from factor analysis.

Table 4.1: Variables, their roles and data type

Variable	Role	Type
Environmental factors (3)	Determinant	Continuous
Unique environmental parameter (1)	Determinant	Continuous
Density of 24 macrobenthic fauna families	Outcomes	Continuous

4.2 Characteristics of the data

4.2.1 Occurrence and abundance of macrobenthic fauna

A total of 24 families were classified into three phyla of Annelida (class Polychaeta), Arthropoda (class Crustacea) and Mollusca (classes Gastropoda and Bivalvia), which comprised the most diverse groups (35.2-98.2% of occurrence).

The Polychaeta was represented by nine families (Capitellidae, Goniadidae, Hesionidae, Nephtyidae, Nereididae, Pilargiidae, Pholoidae, Spionidae and Terebellidae). The Crustacea was also represented by nine families (Aoridae, Isaeidae, Melitidae, Oedicerotidae, Apseudidae, Pseudotanaidae, 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).

Nereididae was the highest occurrence family over all station and month combinations with 98.2% of occurrence, whereas the families of Terebellidae and Stenothyridae had the lowest occurrence (35.2%). Apseudidae was the most abundant family with average density of 40,083.6 ind m^{-2}, while Alpheidae was the least abundant, with average density of 98.2 ind m^{-2}.

The taxonomy, percentages of coverage and densities in ind m^{-2} of the 24 families of macrobenthic fauna in the Middle Songkhla Lake from April 1998 to February 1999 are shown as Table 4.2.

Table 4.2: The taxonomy, percentages of occurrence (%occ) and densities (ind m^{-2}) of the 24 macrobenthic fauna families in the Middle Songkhla Lake from April 1998 to February 1999. [* note the unidentified species in Bivalvia]

Phylum	Class	Order	Family	%occ	Density
Annelida	Polychaeta	Capitellida	Capitellidae	87.0	1,227.3
	Polychaeta	Phyllodocida	Goniadidae	37.0	443.6
	Polychaeta	Phyllodocida	Hesionidae	55.6	698.2
	Polychaeta	Phyllodocida	Nephtyidae	87.0	2,218.2
	Polychaeta	Phyllodocida	Nereididae	98.2	8,507.3
	Polychaeta	Phyllodocida	Pilargiidae	70.4	1,625.5
	Polychaeta	Phyllodocida	Pholoidae	59.3	658.2
	Polychaeta	Spionida	Spionidae	92.6	5,056.4
	Polychaeta	Terebellida	Terebellidae	35.2	1,136.4
Arthropoda	Crustacea	Amphipoda	Aoridae	59.3	2,421.8
	Crustacea	Amphipoda	Isaeidae	87.0	6,900.0
	Crustacea	Amphipoda	Melitidae	94.4	4,438.2
	Crustacea	Amphipoda	Oedicerotidae	66.7	667.3
	Crustacea	Tanaidacea	Apseudidae	90.7	40,083.6
	Crustacea	Tanaidacea	Pseudotanaidae	37.0	4,265.5
	Crustacea	Isopoda	Anthuridae	75.9	3,816.4
	Crustacea	Isopoda	Cirolanidae	37.0	427.3
	Crustacea	Decapoda	Alpheidae	40.7	98.2
Mollusca	Gastropoda	Neogastropoda	Marginellidae	85.2	3,963.6
	Gastropoda	Cephalaspidea	Retusidae	55.6	5,536.4
	Gastropoda	Mesogastropoda	Skeneopsidae	38.9	956.4
	Gastropoda	Mesogastropoda	Stenothyridae	35.2	581.8
	Bivalvia	Unidentified	Unidentified*	44.4	338.2
	Bivalvia	Veneroida	Tellinidae	81.5	17,134.5

4.2.2 Environmental variables

Abiotic data (wDep: water depth, wTemp: water temperature, Sal: salinity, spH: sediment pH, wpH: water pH, DO: dissolved oxygen, TSS: total suspended solids, TN: total nitrogen contents, OC: organic carbon contents, and soil structure as sand, silt, and clay) were used as environmental variables. Their values varied substantially between site and period.

Figure 4.1 plots the water characteristics in the Middle Songkhla Lake from April 1998 to February 1999. The water depth varied to a lesser extent, varying with location from an average of less than 1 m at stations four and nine to more than 2 m at station eight. It was also higher during the rainy season. The water temperature showed decreased values in the rainy season, with a range of 27-34 °C. The salinity increased from close to zero during the rainy season (December to February) to an average close to 20 in the other months. The amount of carbonic acid in rainfall caused the pH of water to decrease, and thus shows the lowest value in December.

Figure 4.2 plots sediment characteristics in the Middle Songkhla Lake from April 1998 to February 1999. The total nitrogen contents at each station was very low (0.02%) during October-February, possibly due to the fact that organic volume is lower during the rainy season since the current sweeps away the organic matter that is spread over the lake. 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 sand, silt and clay percentages sum to 100%.

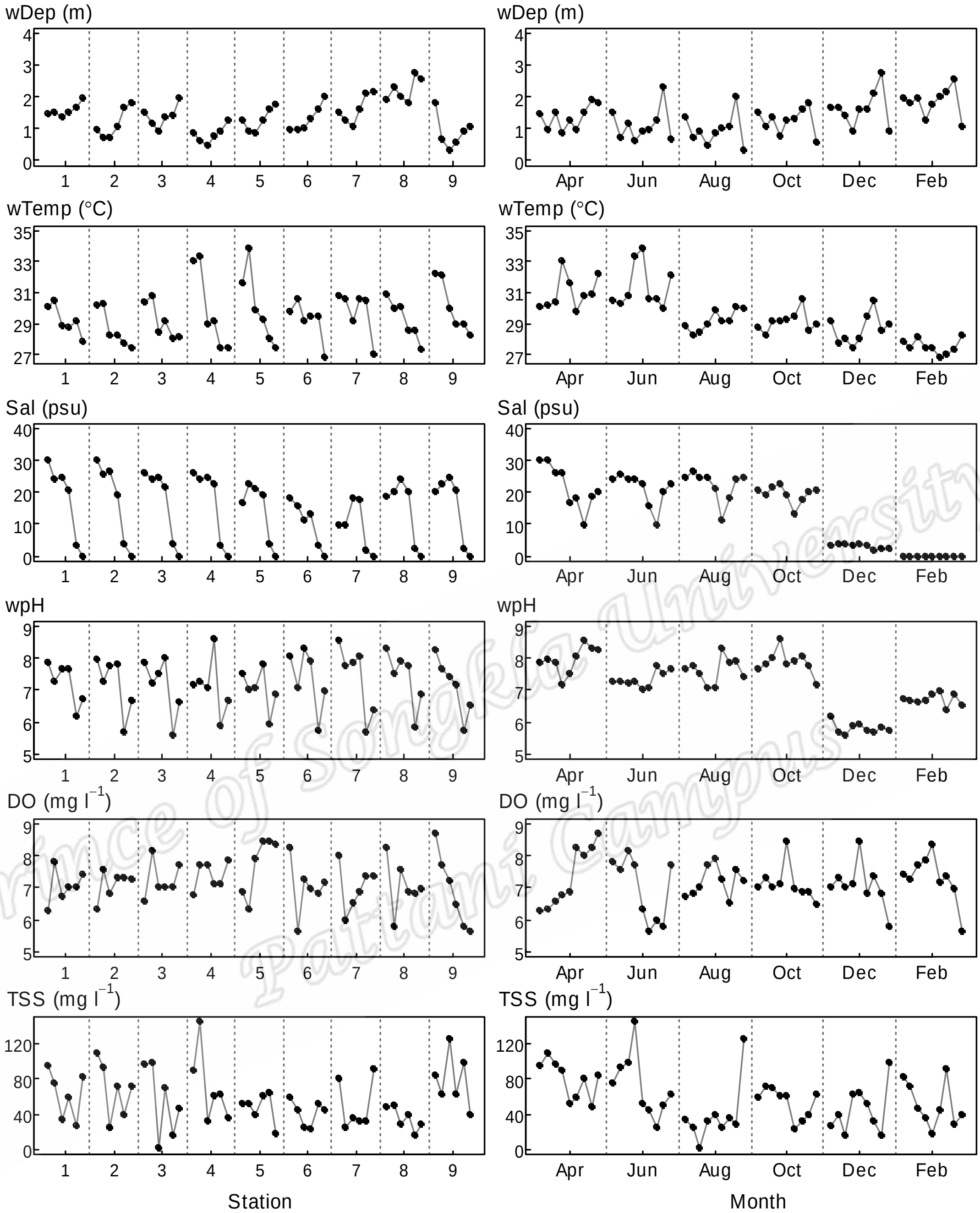

Figure 4.1: The water characteristics: water depth (wDep), water temperature (wTemp), salinity (Sal), water pH (wpH), dissolved oxygen (DO), total suspended solids (TSS) in the Middle Songkhla Lake from April 1998 to February 1999 by station (left panel) and month (right panel)

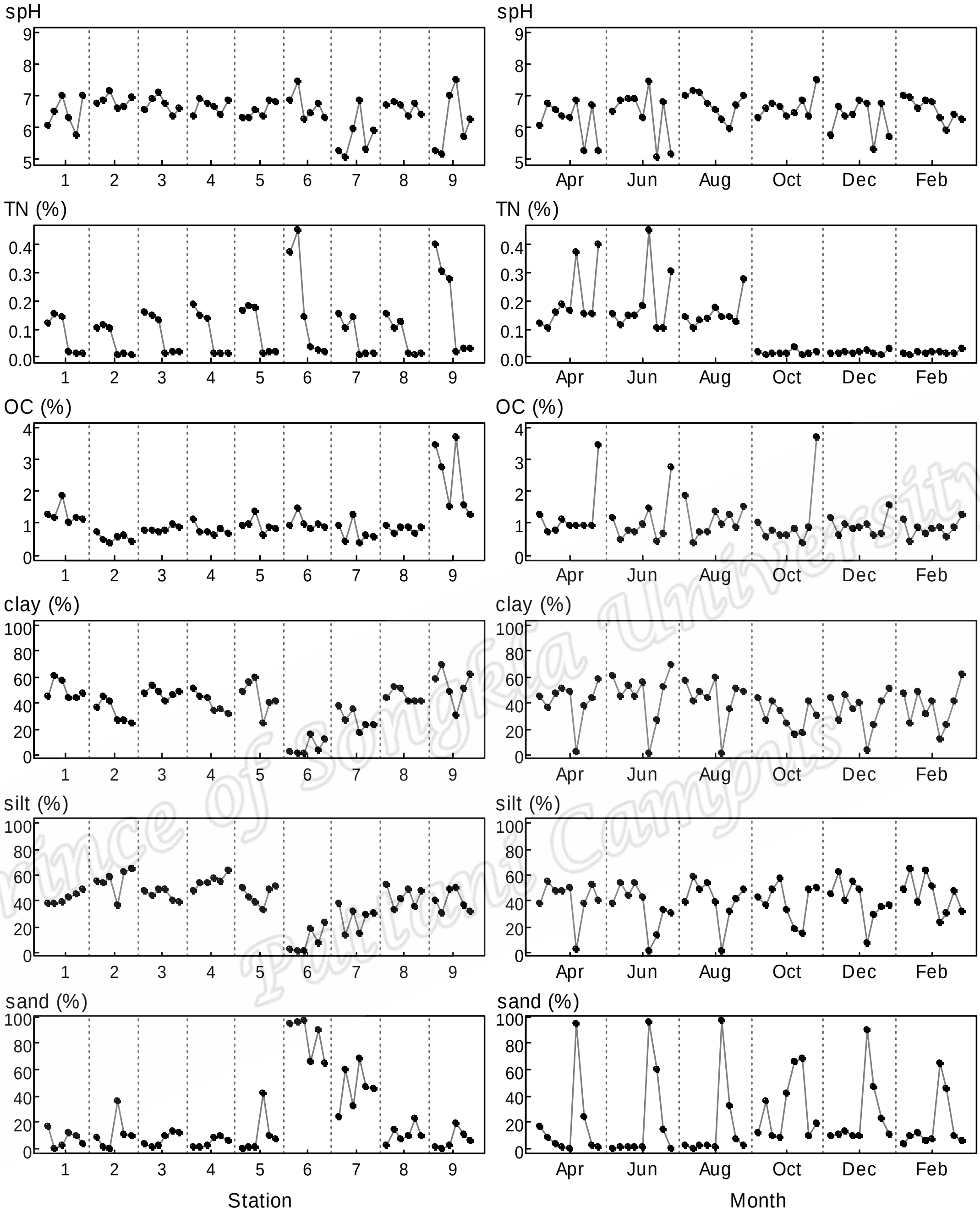

Figure 4.2: The sediment characteristics: sediment pH (spH), total nitrogen contents (TN), organic carbon contents (OC), percentage of clay, silt, and sand in the Middle Songkhla Lake from April 1998 to February 1999 by station (left panel) and month (right panel)

4.3 Factor analysis of environmental variables

Environmental variables were defined as environmental factors based on factor analysis. Factor loadings reflect the correlations between the variables and the extracted factors. Factor loadings of greater than 0.45 are considered significant in this study (Table 4.3). Based on the chi-square test, a factor model with three factors (chi-squared = 20.23, 12 df, *p*-value = 0.06) fitted the data well. These factors do not contain any overlapping variables. 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%.

Table 4.3: Factors and factor loadings for water quality parameters

Variable	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

Interpreting the results, Factor 1 encompasses salinity, containing positive loadings for salinity (Sal) and water pH (wpH), and a negative loading for water depth (wDep) as expected, with deeper water during the rainy season. Factor 2 represents the effect of sediment characteristics in 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 total nitrogen (TN), organic carbon (OC), and water temperature (wTemp); and a negative loading for sediment pH (spH). Each factor was defined as follows:

$$\text{Factor}\,1 = -0.53\times\text{wDep} + 0.70\times\text{wpH} + 0.99\times\text{Sal}\,;$$

$$\text{Factor}\,2 = 0.94\times\text{Sand} - 0.95\times\text{Clay}\,;$$

$$\text{Factor}\,3 = 0.47\times\text{O}C + 0.58\times\text{T}N - 0.57\times spH + 0.54\times wTemp\,.$$

The three factors were included in the MMR model as predictors together with the two singleton variables omitted from the factor analysis (total suspended solids (TSS) and dissolved oxygen (DO)), with each of these five predictor variables scaled to have mean 0 and standard deviation 1.

4.4 Multivariate multiple regression model

The density was taken as log (1 + c × density) with the multiplier c chosen to approximate normality of the error distribution. The choice c = 100 gave residuals satisfying the normality assumption. A total of 24 families of macrobenthic fauna, which were the most diverse groups (> 35% of occurrence) over 54 station-month combinations, were selected. The densities of the selected families were used as the outcome of interest for model fitting.

The environmental factors together with the two singleton variables omitted from the factor analysis (TSS and DO) were first fitted to MMR model. The adequacy of the model can be tested by the MANOVA approach. The model-fitting resulted in the omission of dissolved oxygen, as shown in Table 4.4.

Figure 4.3 is a plot of residuals against corresponding normal quantiles. The plot shows the residuals in the y axis against normal quantiles in the x axis. It was used for checking the normality assumption and the adequacy of model fit to the data, and demonstrates that the points in the residual plot are randomly dispersed around the straight line, thus satisfying the normality assumption, and confirming the appropriateness of the model.

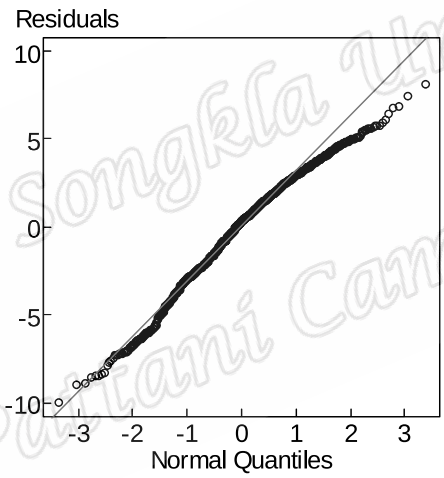

Figure 4.3: Residuals plotted against corresponding normal quantiles for the model

Table 4.4 shows the corresponding individual regression coefficients and standard errors and r-squared values for each family after fitting the MMR model with all four environmental predictors included. The coefficients listed are the ones statistically significant at 5% and 1% (in bold).

Since there are 96 regression coefficients in all, and 5% of these would be expected to have *p*-values < 0.05 even if all their corresponding parameters was zero, the five largest *p*-values < 0.05 are italicized to indicate failure to “honest” significance.

Table 4.4: Regression coefficients and standard errors (in parenthesis) from fitting MMR model with the four environmental predictors; coefficients with *p*-values < 0.05 are shown; those adjudged not honestly statistically significant are shown in italics and those with *p*-values < 0.01 are shown in bold; the coefficients that were not statistically significant are labeled "*ns*". [* note the unidentified species in Bivalvia]

Family	Intercept	Factor 1	Factor 2	Factor 3	TSS	r^2
Capitellidae	6.27(0.37)	0.16(0.23)^{ns}	0.06(0.20)^{ns}	–0.20(0.28)^{ns}	–0.47(0.41)^{ns}	0.05
Goniadidae	2.59(0.41)	**1.01**(0.25)	0.04(0.22)^{ns}	–0.76(0.31)	–1.10(0.45)	0.31
Hesionidae	3.87(0.45)	**0.76**(0.28)	–0.45(0.24)^{ns}	0.11(0.34)^{ns}	–0.85(0.49)^{ns}	0.23
Nephtyidae	6.58(0.38)	–0.16(0.23)^{ns}	–0.37(0.21)^{ns}	0.25(0.28)^{ns}	0.35(0.42)^{ns}	0.11
Nereididae	8.44(0.25)	0.28(0.15)^{ns}	0.30(0.13)	0.42(0.19)	0.01(0.27)^{ns}	0.25
Pilargiidae	5.17(0.47)	0.36(0.29)^{ns}	**–0.68**(0.25)	–0.18(0.35)^{ns}	–0.06(0.51)^{ns}	0.17
Pholoidae	3.97(0.41)	–0.61(0.25)	0.43(0.22)^{ns}	–0.54(0.30)^{ns}	–0.05(0.44)^{ns}	0.32
Spionidae	7.38(0.35)	*0.45*(0.21)	0.08(0.19)^{ns}	0.17(0.26)^{ns}	–0.51(0.38)^{ns}	0.12
Terebellidae	2.48(0.37)	0.20(0.23)^{ns}	**1.04**(0.20)	*0.60*(0.28)	–0.91(0.40)	0.44
Aoridae	4.48(0.48)	0.00(0.30)^{ns}	**0.88**(0.26)	0.54(0.36)^{ns}	*1.08*(0.53)	0.25
Isaeidae	7.16(0.42)	0.13(0.26)^{ns}	*0.47*(0.23)	0.54(0.31)^{ns}	–0.42(0.46)^{ns}	0.15
Melitidae	7.68(0.30)	0.20(0.18)^{ns}	–0.01(0.16)^{ns}	0.58(0.22)	0.24(0.32)^{ns}	0.24
Oedicerotidae	4.62(0.47)	–0.11(0.29)^{ns}	–0.36(0.25)^{ns}	–0.34(0.35)^{ns}	–0.42(0.51)^{ns}	0.08
Apseudidae	8.91(0.49)	–0.26(0.30)^{ns}	–0.31(0.26)^{ns}	–0.12(0.36)^{ns}	0.10(0.53)^{ns}	0.05
Pseudotanaidae	2.79(0.45)	–0.64(0.28)	**0.81**(0.24)	0.27(0.34)^{ns}	–0.63(0.49)^{ns}	0.34
Anthuridae	5.58(0.45)	0.11(0.28)^{ns}	**0.69**(0.24)	–0.40(0.34)^{ns}	–0.15(0.49)^{ns}	0.19
Cirolanidae	2.38(0.39)	0.42(0.24)^{ns}	0.51(0.21)	0.67(0.29)	0.46(0.42)^{ns}	0.29
Alpheidae	2.39(0.40)	–0.19(0.25)^{ns}	–0.23(0.22)^{ns}	0.21(0.30)^{ns}	0.46(0.44)^{ns}	0.07
Marginellidae	7.02(0.41)	–0.38(0.25)^{ns}	0.42(0.22)^{ns}	0.11(0.31)^{ns}	–0.33(0.45)^{ns}	0.16
Retusidae	4.32(0.52)	–0.16(0.32)^{ns}	–0.31(0.28)^{ns}	–0.42(0.39)^{ns}	**–1.59**(0.57)	0.21
Skeneopsidae	2.58(0.48)	0.20(0.29)^{ns}	–0.14(0.26)^{ns}	–0.05(0.35)^{ns}	–0.28(0.52)^{ns}	0.02
Stenothyridae	2.32(0.40)	0.35(0.25)^{ns}	–0.42(0.22)^{ns}	0.65(0.30)	–0.26(0.44)^{ns}	0.25
Unidentified*	2.98(0.44)	0.71(0.27)	0.04(0.24)^{ns}	–0.10(0.33)^{ns}	*–0.99*(0.48)	0.16
Tellinidae	7.02(0.52)	0.57(0.32)^{ns}	–0.17(0.28)^{ns}	0.10(0.39)^{ns}	–0.39(0.56)^{ns}	0.09

The densities of Hesionidae, Spionidae, unidentified species in Bivalvia, and Goniadidae were positively related with Factor 1 (mainly salinity) while Pholoidae and Pseudotanaidae were negatively related with this Factor 1.

Eight families (Pseudotanaidae, Pilargiidae, Isaeidae, Anthuridae, Nereididae, Cirolanidae, Terebellidae, and Aoridae) were associated with Factor 2 (sand-clay excess), only Pilargiidae related to the clay habitat.

Six families (Goniadidae, Nereididae, Cirolanidae, Terebellidae, Melitidae, and Stenothyridae) were associated with Factor 3 (Physico-chemical properties); only Goniadidae was negatively related with Factor 3.

Five families (unidentified species in Bivalvia, Goniadidae, Terebellidae, Aoridae, and Retusidae) were associated with total suspended solids (TSS), only Aoridae being positively related with TSS.

Finally, eight families (Capitellidae, Nephtyidae, Oedicerotidae, Apseudidae, Alpheidae, Marginellidae, Skeneopsidae, and Tellinidae) showed no evidence of any environmental predictors.

Goniadidae and Terebellidae could be predicted by three environmental factors (r-squared statistics of 31% and 44%, respectively). Nereididae, Aoridae, Pseudotanaidae, Cirolanidae, and unidentified species in Bivalvia could be predicted by two factors, having r-squared values ranging from 16% to 34%. Hesionidae, Pilargiidae, Pholoidae, Spionidae, Isaeidae, Melitidae, Anthuridae, Stenothyridae, and Retusidae could be predicted by one factor, having r-squared values ranging from 12% to 32%.

The observed densities and fitted values from model fitting are plotted by the periods (Figure 4.4) and by station (Figure 4.5). Capitellidae, Nephtyidae, Oedicerotidae, Apseudidae, Marginellidae, and Tellinidae could be widespread for each observed period and station. Whereas, Alpheidae and Skeneopsidae could not be predicted because the data was underestimate.

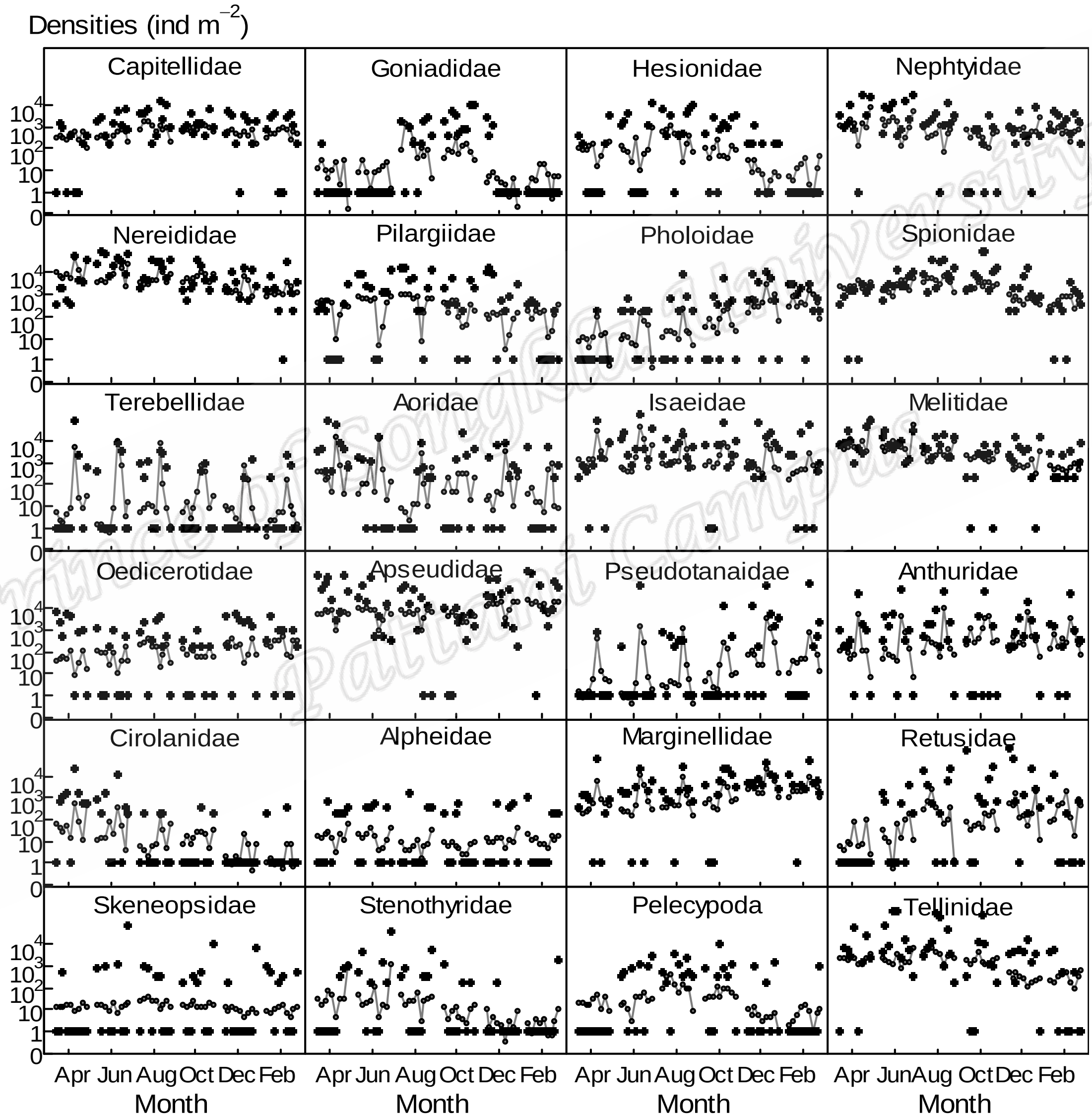

Figure 4.4: Plots of the observed densities (dots) and the predicted densities (dots with model fitted line) for the 24 families by the periods of observation

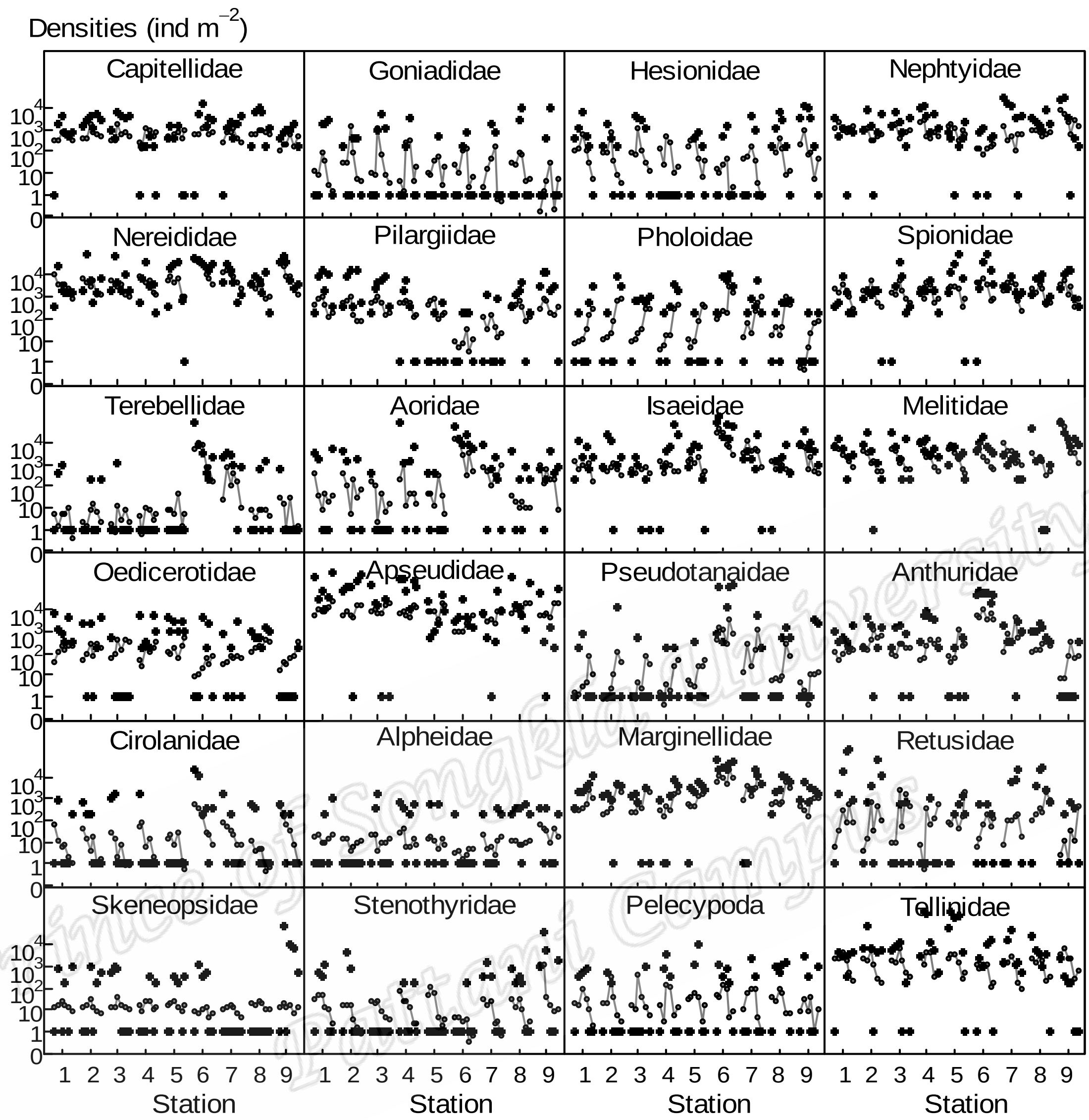

Figure 4.5: Plots of the observed densities (dots) and the predicted densities (dots with model fitted line) for the 24 families by sampling stations