

Chapter 2

Methodology

This chapter describes the methods used in the study. These methods include the following components.

- (a) the study design;
- (b) measurement procedures;
- (c) the data management;
- (d) computer program; and
- (e) the methods used for the statistical analysis.

2.1 Study Design

A cross-sectional study design is used, based on (a) depth (water elevation) and (b) salinity collected by two boats at different locations on 22 occasions in Pattani Bay mouth. Measurements were taken from each side of two boats. The data on salinity and the depth were collected in the middle of Pattani Bay approximately 5 kilometres from the Rusamilae jetty in the north-east direction. Each boat had 11 occasions to sample the water and measure the water depth in the Pattani Bay mouth. To obtain the salinity and water depth, the teamwork was divided into two groups working as follows:

1. A navigator using GPS;
2. Two water depth measurers on each boat side (about 2 meters width) using a water depth recorder; and
3. Two water sample collectors on each boat side for salinity measurement.

The boats were supposed to proceed together in a north-east direction for approximately 5 kilometres from the starting point at Rusamilae jetty. Each boat then had seven locations and 11 occasions to sample water and measure water depth, as shown in Figure 2.1. Boat 2 was to remain North of Boat 1 and stop close to GPS position $6^{\circ}55.18$ N and $101^{\circ} 15.75$ E. Boat 1 was to remain South of Boat 2 and stop close to GPS position $6^{\circ} 54.75$ and $101^{\circ}16.0$ E, as shown in Table 2.1. The planned

GPS locations for each successive data collection occasion for each boat were as shown in Figure 2.2.

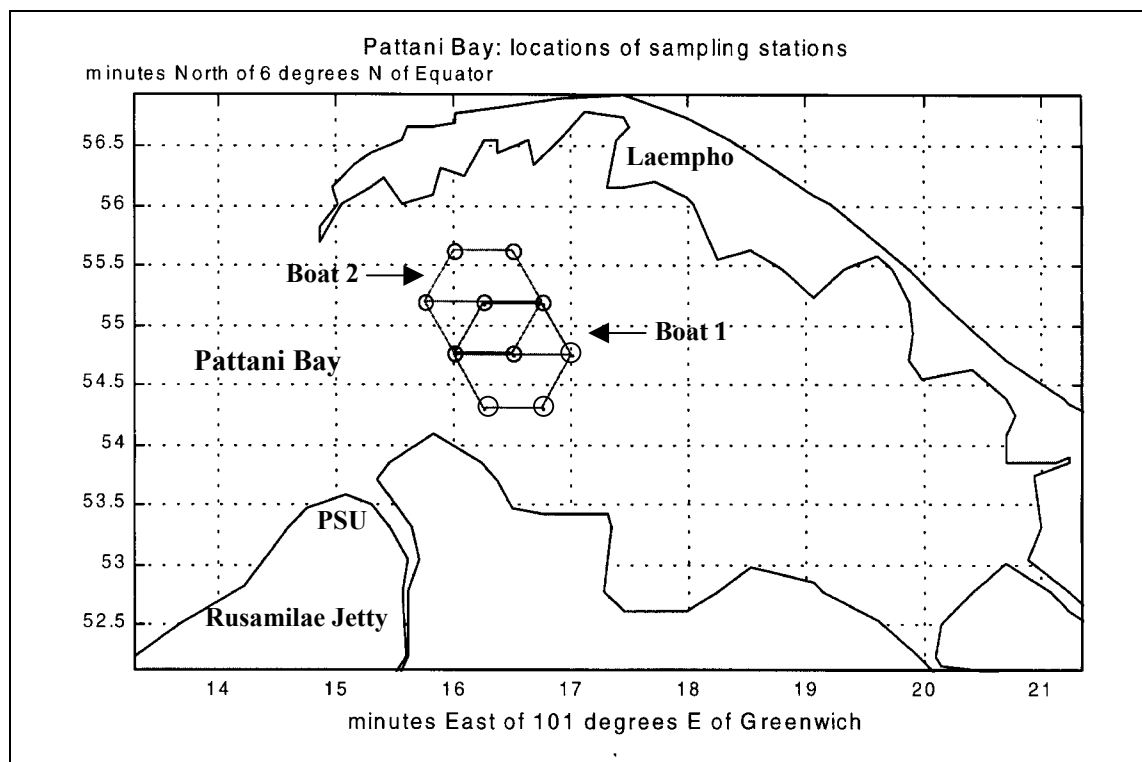


Figure 2.1: Map of Pattani Bay showing the sampling location

Occasion	Location	
	Boat 1	Boat 2
1	6° 54.75' N, 101° 16.00' E	6° 55.18' N, 101° 15.75' E
2	6° 54.75' N, 101° 16.50' E	6° 55.18' N, 101° 16.25' E
3	6° 54.75' N, 101° 17.00' E	6° 55.18' N, 101° 16.75' E
4	6° 55.18' N, 101° 16.75' E	6° 55.61' N, 101° 16.50' E
5	6° 55.18' N, 101° 16.25' E	6° 55.61' N, 101° 16.00' E
6 (=1)	6° 54.75' N, 101° 16.00' E	6° 55.18' N, 101° 15.75' E
7 (=2)	6° 54.75' N, 101° 16.50' E	6° 55.18' N, 101° 16.25' E
8 (=3)	6° 54.75' N, 101° 17.00' E	6° 55.18' N, 101° 16.75' E
9	6° 54.32' N, 101° 16.75' E	6° 54.75' N, 101° 16.50' E
10	6° 54.32' N, 101° 16.25' E	6° 54.75' N, 101° 16.00' E
11 (=1)	6° 54.75' N, 101° 16.00' E	6° 55.18' N, 101° 15.75' E

Table 2.1: The sampling locations using GPS

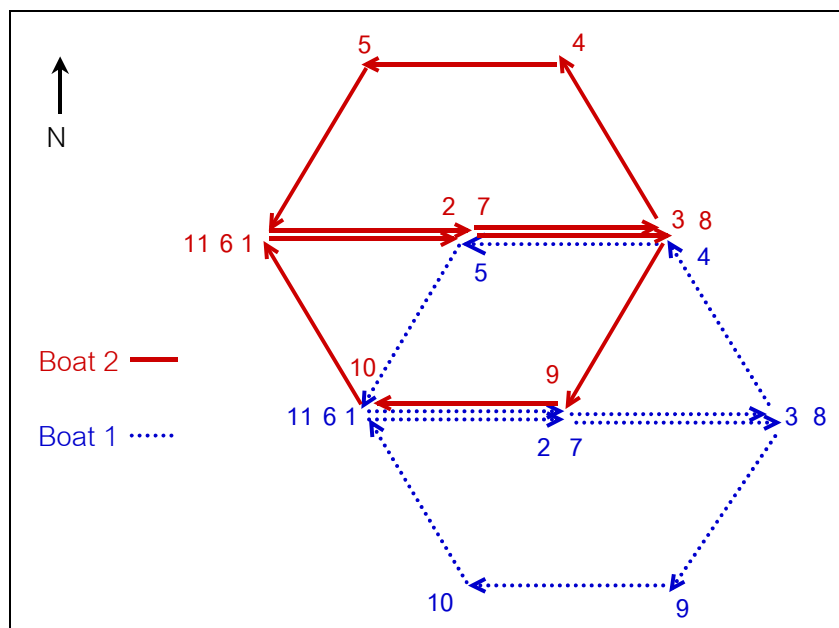


Figure 2.2: The data collection design for the two boats

2.2 Measurement Procedures

Materials and equipment

1. A fishing line of about 10 meter's length
2. A bottle attached to a bamboo stick and a cork
3. GPS
4. Film containers
5. Icebox
6. Three Salinometers (ATAGO, Japan)

Each depth measurer had a fishing line of about 10 meter's length for two depth measurements by slowly dropping the sinker line until it reached the bottom, and marking the water-level position on the line measuring the distance and recording it in a log book corresponding to each location.

Each salinity measurer had a bottle attached to a bamboo stick and a cork for collecting water samples. The water samples were taken at 50 cm below the surface with the bottle and poured into two film containers to about $\frac{3}{4}$ capacity, sealed, labeled, and then placed in the icebox for later salinity measurements.

In the laboratory three salinometers were calibrated using distilled water (salinity = 0) before measuring the salinity. Each of the 44 water samples was to be read by six readers as follows: 1: Montri, 2: Sunari, 3: Thitiphit, 4: Siriporn, 5: Chaiya and 6: Zainab.

Eventuality

At 8 am on 30 July 1999, the research team assembled at the Rusamilae jetty and proceeded to the Bay to make the measurements. However, Boat 2 did not exactly follow the experimental design plan. The actual GPS positions corresponding to the 11 occasions for each boat were as shown in Table 2.2. The actual locations are shown in Figure 2.3.

Boat 1			Boat 2		
occasion	location		occasion	location	
	Latitude(N)	Longitude(E)		Latitude(N)	Longitude(E)
1	6° 54.79'	101° 15.99'	1	6° 55.20'	101° 15.75'
2	6° 54.76'	101° 16.53'	2	6° 55.19'	101° 16.29'
3	6° 54.78'	101° 17.01'	3	6° 55.19'	101° 16.73'
4	6° 55.21'	101° 16.77'	4	6° 55.62'	101° 16.52'
5	6° 55.24'	101° 16.21'	5(=4)	6° 55.61'	101° 16.53'
6(=1)	6° 54.74'	101° 16.04'	6(=1)	6° 55.19'	101° 15.76'
7(=2)	6° 54.70'	101° 16.48'	7(=2)	6° 55.17'	101° 16.28'
8(=3)	6° 54.75'	101° 17.00'	8	6° 54.70'	101° 16.50'
9	6° 54.24'	101° 16.51'	9(=8)	6° 54.74'	101° 16.51'
10	6° 54.31'	101° 16.21'	10	6° 54.76'	101° 16.00'
11(=1)	6° 54.77'	101° 16.02'	11(=1)	6° 55.20'	101° 15.74'

Table 2.2: The sampling location using GPS (in practice)

The locations that differed between the initial design for Boat 2 occurred on occasions 5 and 8, from (6° 55.61' N, 101° 16.0' E) to (6° 55.61' N, 101° 16.53' E) and from (6° 55.18' N, 101° 16.75' E) to (6° 54.70' N, 101° 16.50' E), respectively. Consequently, we collected data at nine locations instead of ten locations.

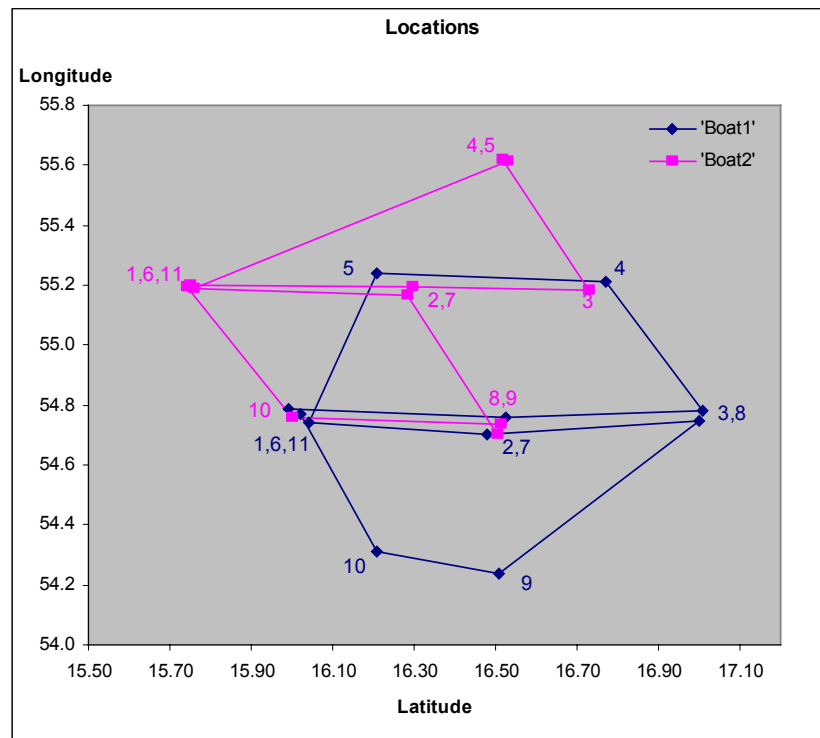


Figure 2.3: The direction on occasion in the two boats

2.3 Data Management

The raw data of salinity and depth were kept in a Microsoft Access database (see Appendix). These data were stored in the database **salinity.mdb**, which had two base tables, **depth** and **salinity**, containing the position, time, observer, latitude, longitude, boat reader, instrument, location and boat. The relationships between salinity and depth data are shown in Figure 2.4.

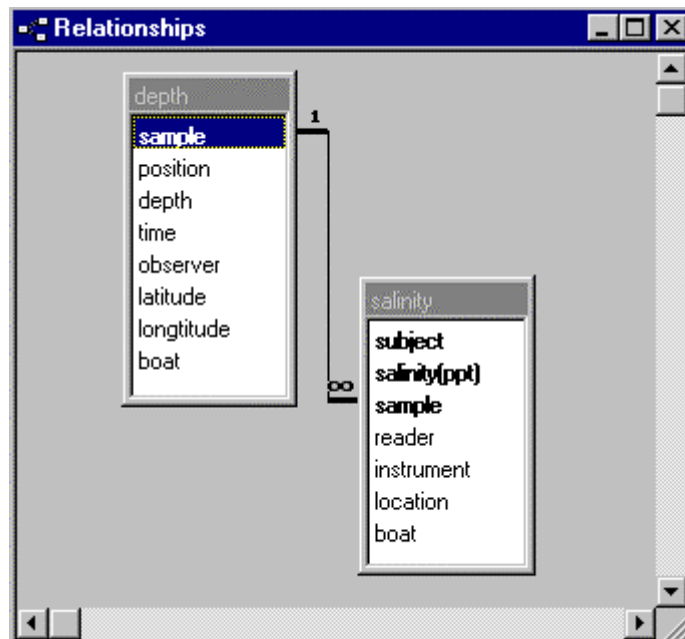


Figure 2.4: Relationship between depth and salinity in database

2.4 Computer programs

The following computer programs were used for data analysis and thesis preparation.

1) Asp

Asp is a suite of function for graphing and analyzing statistical data. These programs run under Matlab Version 5 (Hanselman & Littlefield, 1997). It was mainly used to perform preliminary data and regression analysis.

2) Microsoft Excel

This program was mainly used to collect the data for this research. Some function is helpful in plotting graph.

3) Microsoft Word

This program was mainly used to write and print the report of this research.

4) EcStat in Microsoft Excel

EcStat is an add-in to Microsoft Excel 97. It is a suite of routines for graphing and analyzing statistical data using an IBM-compatible PC.

EcStat is being developed because there is need for both specialist and non-specialists in Statistics to be able to analyze data easily and to display the relevant results in a concise, appropriate way showing the information content of the data. *EcStat* was mainly used in 'Comparison' and 'Relation' commands. The results of 'Comparison' is a one-way analysis of variance (anova), and the result of 'Relation' is a graph with fitted linear relation shown on the plot.

2.5 Statistical Methods

Numerical and graphical summaries of variables, and comparisons based on two-sample t-test, one-way and two-way analysis of variance were used in the preliminary data analysis. Correlation coefficient and simple statistical models were used, including multiple regression analysis to develop the model of salinity for prediction of the salinity in Pattani Bay.

1. Two-sample t-test

The two-sample t-test that the population means are the same, and the statistic is obtained as follows (McNeil, 1996 : 44)

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

when s is the *pooled* sample standard deviation, defined by subtracting the sample mean from each sample to give a set of $n_1 + n_2$ residuals, dividing the sum of the squares of these residuals by $n_1 + n_2 - 2$, and taking the square root of the result. If s_1 and s_2 denote the standard deviations of the two samples, respectively, it may be shown that the pooled sample standard deviation is given by the formula

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

A p-value is now obtained from the table of the two-tailed t distribution with $n_1 + n_2 - 2$ degrees of freedom.

2. One-way Analysis of Variance

One-way Analysis of Variance is the method used for the analysis of data in which the outcome is continuous and determinant is categorical. To assess the statistical significance of the comparison, an F -statistic is compared with an appropriate distribution to get a p-value for testing a null hypothesis. This null hypothesis states that the population means of the continuous outcome variable (depth or salinity in this case) are the same for each level of the determinant (boat or occasion or reader). The F -statistic is defined as follows (McNeil, 1996 : 67):

$$F = \frac{(S_0 - S_1)/(c-1)}{S_1/(n-c)}$$

where c is the number of categories of the determinant and n is the number of observations. In this formula S_0 is the sum of squares of the outcomes after subtracting their overall mean, and S_1 is the corresponding sum of squares after subtracting different means for each level of the determinant. The aim of one-way analysis is to compare the mean (depth or salinity) for each boat and occasion. The p-value is the tail area of the F distribution with $c-1$ and $n-c$ degrees of freedom.

3. Two-way Analysis of Variance

Two-way analysis of variance is the extension of one-way analysis of variance used to compare the means of a continuous outcome variable for different values of a categorical determinant, after adjusting for the effect of a categorical covariate. The p-value is based on an F -statistic defined as (McNeil, 1996 : 73)

$$F = \frac{(S_2 - S_{12})/(c-1)}{S_{12}/(n-c-r+1)}$$

where r is the number of levels of the covariate. In this formula S_2 is the sum of squares of the outcomes after subtracting different means for each level of the covariate, and S_{12} is the corresponding sum of squares after subtracting different means for each additive combination of levels of the determinant and the covariate. The p-value is the tail area of the F distribution with $c-1$ and $n-c-r+1$ degrees of freedom.

These formulas assume that there is exactly one observation for each combination of levels of the determinant and the covariate. If this assumption is not met, some adjustment is needed, and the E-M algorithm can be used (Dempster et al, 1977 : 1-38).

The assumptions required for analysis of variance are as follows:

- (1) the errors have constant variation;
- (2) the errors are independent and normally distributed.

If assumption (1) is not satisfied it may be necessary to transform the data before the analysis. If assumption (2) is not met it may be necessary to remove outliers.

4. Correlation Coefficient

The correlation coefficient is a measure of the linear or straight-line, relationship between variables and level of relation. The model of correlation coefficient is defined as (McNeil, 1998 : 181)

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

It shows that r ranges from a minimum of -1 to maximum value of 1. A correlation coefficient equal to 0 indicates no linear relationship between the two variables.

5. Regression Analysis

The multiple linear regression model is

$$Y = \beta_0 + \sum \beta_i x_i + \varepsilon,$$

where Y is the outcome variable, β_0 is a constant, $\{\beta_i\}$ is a set of parameters ($i = 1$ to p), and x_i is a set of determinants ($i = 1$ to p).

The model is fitted to data using least squares, which minimizes the sum of squares of the residuals.

Multiple linear regression analysis is used to find the relationship between salinity with depth, time, latitude and longitude, giving estimated values b_0, b_1, \dots, b_p for the coefficients. There are three assumptions for linear regression analysis, namely

- (1) the relation is linear;
- (2) the variability of the errors (in the outcome variable) is uniform, and
- (3) these errors are normally distributed.

If assumption (1) and/or (2) are not met, a transformation of the data may be appropriate (McNeil, 1998 : 185).