Chapter 2

Methodology

This chapter describes the data sources and statistical methods used to analyze these data.

2.1 Data sources and measurements

The catch of blue swimming crab data in Songkhla Lake were obtained from National Institute of Coastal Aquaculture (NICA), Songkla province from January 1, 2003 to December 31, 2006. The monthly data were collected using three main gear types. They were trap, set bag net and gill net.

The trap is a small sized stationary fishing gear, combined with seine wing-like barriers used for re-directing and then trapping crab, shrimp and fish as shown in Figure 2.1. Trap selectivity relies on the fish moving actively into the trap and catches depend on the duration of the soak, i.e. the time the trap is left on the fishing ground.

Figure 2.1: The main gear types used in Songkhla Lake: trap

The set bag net is a traditional fishing gear widely used in the coastal areas of Asia. It is usually place in a relatively deep channel with a relatively strong current. Having a small mesh size, it is non selective for size and species as shown in Figure 2.2.

Figure 2.2: The main gear types used in Songkhla Lake: set bag net

The gill net is usually long rectangular net set vertically, where the upper edge has floats while the foot rope has sinkers. Gill nets (both surface net and drift types) are passive gears that the fish have to swim into as shown in Figure 2.3.
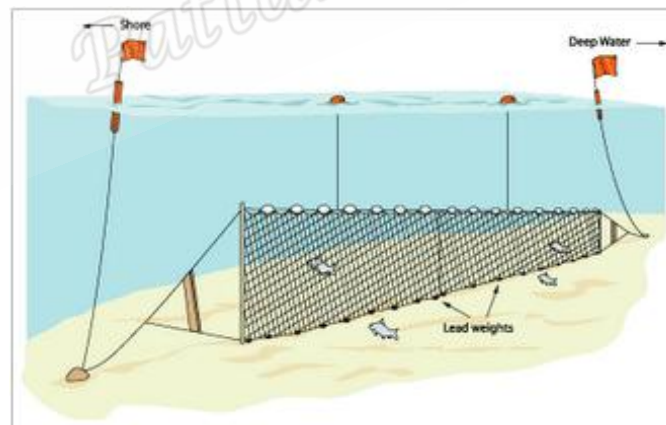

Figure 2.3: The main gear types used in Songkhla Lake: gill net

Some environmental factors collected from Southeastern Regional Meteorological Center, Songkla province during 2003–2006. Daily data were used to calculate the mean for monthly data. The environmental factors are air temperatures, morning air

temperature, rainfall, relative humidity and water level. Data of water level was hourly measured by Hydrographic Department Royal Thai Navy (Hydrographic Department Royal Thai Navy, 2003-2006). The monthly values were calculated using the mean.

Air temperatures comprise morning air temperature and average air temperature. The air temperature measured every three hours from 1 a.m. to 12 a.m. on the next day, and calculated daily average air temperatures. The morning air temperature measured at 7 a.m. every day.

Rainfall is the volume of rainfall measured at four stations in Songkla province. They are Muang district, Agricultural Station Kho Hong; Had Yai district and Sadao district. The rainfall was measured from 7 p.m. to 7 p.m. on the next day.

Humidity is the amount of water vapor in the air. Relative humidity is defined as the ratio of the partial pressure of water vapor in a parcel of air to the saturated vapor pressure of water vapor at a prescribed temperature. Relative humidity is an important metric used in forecasting weather. Humidity indicates the likelihood of precipitation, dew or fog.

Water level is the heights of water predicted in meters above the average sea level. This measured hourly at mouth of Songkhla Lake that connect to the Gulf of Thailand (Latitude 07^0 12' 58" N and Longitude 100^0 34' 50" E). They were measured by Hydrographic Department Royal Thai Navy (Hydrographic Department Royal Thai Navy, 2003-2006) as shown in Figure 2.4.

Figure 2.4: Locations of water level measurements in the south of Thailand

2.2 Path diagram

The path diagram of this study was shown in Figure 2.5. The outcome was of blue swimming crab catch weight in kilogram. The determinant variables comprised year, month, gear and environmental factors (air temperatures, morning air temperature, rainfall, relative humidity and water level).
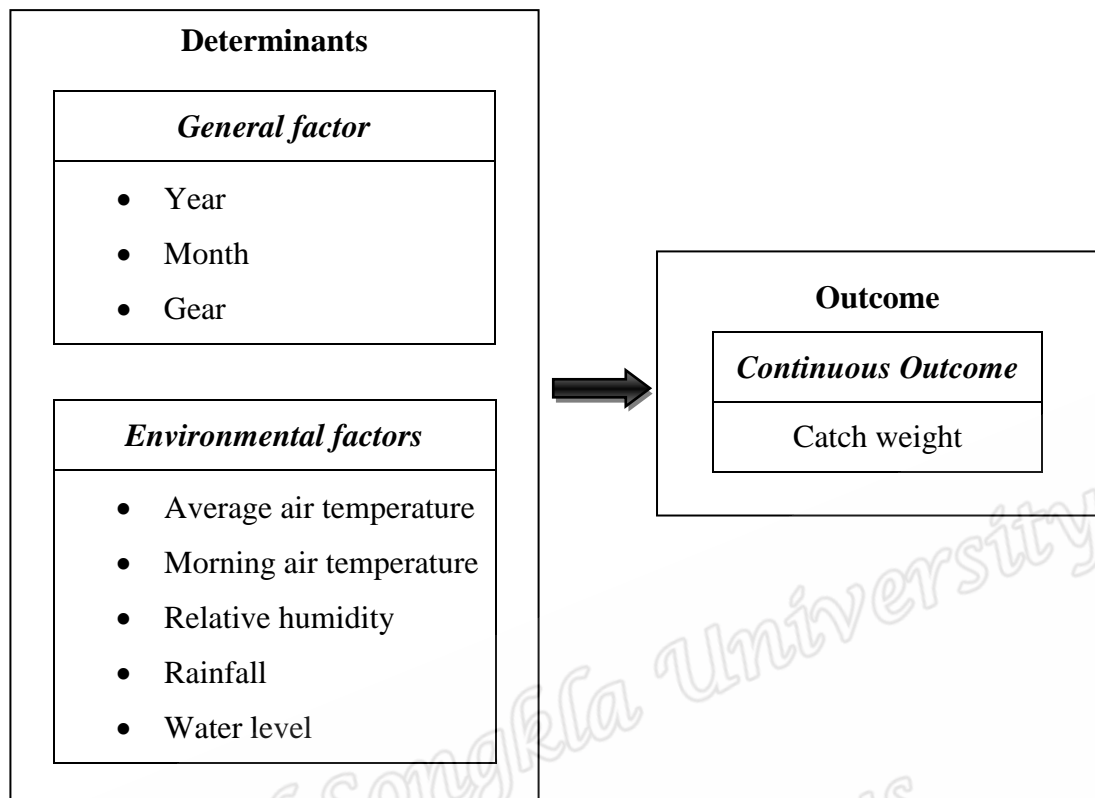
Figure 2.5: Path diagram of the study

2.3 Data analysis

In preliminary data analyses, we used time series plot for the total catch weight and

for the catch weight classified by gear types. Then we plot histogram of the catch and

the catch by three gear types. To satisfy statistical assumptions of normality, the

monthly catch weights were transformed by using natural logarithms to reduce the

skewness. We used box plot to compare the natural logarithm of catch weight with

different gears, years and month. For environmental factors we showed descriptive

statistics and time series plot of each factor. We examined the relationship between

each environmental factor with the natural logarithm of catch weight.

2.4 Statistical methods

Coefficient of determination

Coefficient of determination or percentage of variance explained or R^2 is a common test for evaluating goodness of fit. It gives the percentage of total variation in the dependent variable explain by the regression line (Faraway, 2002). The range is $0 \leq R^2 \leq 1$, a value closer to 1 indicating better fits. For linear regression the coefficient of determination (R^2) is the square of the correlation coefficient (r^2) where r is the correlation coefficient between x and y. An equivalent definition is

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \qquad (2.1)$$

Regression Analysis

Simple linear regression is used to analyze data in which both the outcome and the determinants are continuous variables. A scatter plot may be used to display the data, and then the slope of a fitted straight line is used to represent the association between the determinant and outcome.

In conventional statistical analysis the lines fitted is *the least squares lines*, which minimize the square of the distances of the points to the line, measured in the vertical direction. This line is also called the *regression line*, and may be represented as

$$y = a + bx \qquad (2.2)$$

where

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad (2.3)$$

and

$$a = \bar{y} - b\bar{x} \qquad\qquad (2.4)$$

Where a is the *intercept* and b is the *slope* or *regression coefficient*. If there is more than one determinant, the method generalized to multiple linear regressions represented as

$$y = \beta_0 + \sum \beta_i x_i + \varepsilon \qquad\qquad (2.5)$$

Where y is the outcome variable, β_0 is a constant, β_i is a set of parameters (i is the number of determinants), and x_i is a set of determinants.

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

There are three assumptions that have to be check when using linear regression analysis. First, the association between outcome and determinant variable is linear. Second, the variability of the error (in the outcome variable) is uniform and these errors are normally distributed. If these assumptions are not met, a transformation of the data may be appropriate. Linear regression analysis may also used when one or more of the determinant are categorical. In this case the categorical determinant is broken down into c-1 separate binary determinants, where c is number of categories. The omitted category is taken as the baseline or referent category (McNeil, 1996). The linear regression model with categorical determinant can also be fit using sum contrast (Venable and Ripley, 2002).

Sum contrast

Sum contrast were used to calculate the standard errors when constructing confidence intervals for comparing means without selecting a reference group and thus give informative confident intervals for comparing the mean for each level of each factor with the overall mean (Tongkumchum and McNeil, 2009).

All graphical and statistical analyses were carried out using R (R Development Core Team, 2008).