CHAPTER 3

METHODOLOGY

3.1 Study Area

In Thailand, the Pollution Control Department (PCD) has responsibility to monitor air quality in general and crisis areas. Currently, PCD operates eight permanent and approximately fifteen temporary stations in Bangkok (PCD 1996) for the purpose of monitoring ambient air quality of the entire city. Locations of eight permanent stations are shown in Fig. 3.1. For this study, a monitoring site at Dindang district was selected. This location was not exactly at the center of Bangkok but it represents a major residential area of the city (BMA 1998).

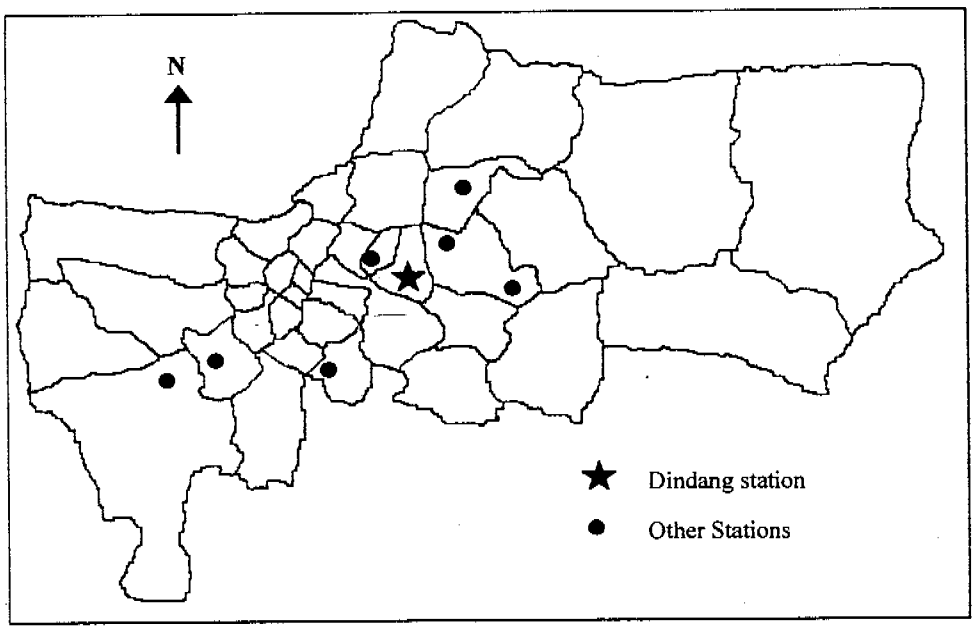


Fig. 3.1 Map of Bangkok area

3.2 Air Pollution Sampling and Analyses

Air quality data at the Dindang sampling site were collected for 730 days from January 1, 1997 to December 31, 1998. Five air quality parameters; SO_2, NO_2, CO, O_3, and PM_{10} (particulate matters with diameter less than 10 μm) were measured and used in this study.

Air samples were drawn through a sampling port located at the top of the monitoring station. Its intake port is 3 meters high from the ground. Sulfur dioxide (SO_2) was measured with a fluorescence SO_2 analyzer (Model 100) from Advanced pollution Instrumentation, API; CA, USA (Range: 0-1000 ppb, Precision 99.5%).

Nitrogen dioxide (NO_2) was measured with an API (Model 200A) chemiluminescence NO_x analyzer (Range: 0-1000 ppb, Precision 99.5%). Carbon monoxide (CO) was measured with an API (Model 300) gas filter correlation CO analyzer (Range: 0-100 ppm, Precision 99.5%). Ozone (O_3) was measured with an API (Model 400) UV absorption O_3 analyzer (Range: 0-1000 ppb, Precision 99.5%). Particulate matters (PM_{10}) was measured with a beta gauge automated particle sample (Range: 0-2000 $\mu g/m^3$, Precision 92%) from Wedding and Associates, CO, USA. The instruments record SO_2, NO_2, CO, O_3 and PM_{10} concentrations analyzed, and transmit the results to a data logger every hour.

3.3 Data Analysis by Factor Analysis

FA is applicable whenever a quantity can be expressed as a linear combination of others. In the present case, we make an assumption of expressing the concentration of each species found in a sample as a linear combination of the concentration of the species found in the source times the contribution of the source to the sample for all possible sources (Rachdawong and Christensen 1997) or;

$$d_{ik} = \sum_{j=1}^{n} c_{ij} r_{jk} \tag{3.1}$$

where;
d_{ik} = the concentration of the compound i in the sample k
c_{ij} = the concentration of the compound i in the source j as represented by factor loadings
r_{jk} = the contribution from source j to sample k as represented by factor scores

The sum is extended over the number of factors or sources, which adequately account for the measurement in question. In matrix notation the equation can be written as;

$$D = CR \tag{3.2}$$

The main goal of FA is to find the factor loading matrix C that is also the source composition matrix. R is the matrix of factor score representing source contributions and D is the matrix of observations with experimental error. Factor analysis tends to find the minimum number of factors which reproduce the experimental data matrix D with acceptable error limit.

3.4 Steps in Factor Analysis

An average of 24-hr data, matrix D, is input into the factor analysis model (consisting of rows representing samples) and columns (representing variables). Data matrix is then subject to the following procedures.

Step 1. The data matrix standardization

All data are transformed to standardized data as in this following relation;

$$d = \frac{x - \bar{x}}{s} \tag{3.3}$$

where;
x is daily data, $\bar{x}$ is mean and s is standard deviation

Step 2. Formation of correlation matrix

A correlation matrix Z is computed as follows;

$$Z = \frac{D^t D}{N} \tag{3.4}$$

where;
Matrix D^t is a tranposed matrix of D

Step 3. Extraction of eigenvalues and eigenvectors

The Jacobi method of factor extraction is applied to the correlation matrix. Subsequently, an orthonormal eigenvector Q (columns of Q are eigenvectors) and an eigenvalue matrix E, whose diagonal elements are eigenvalues λ, are obtained. The relations among Q, E, and Z can be shown to be;

$$Z = QE^tQ \tag{3.5}$$

Step 4. Factor loading and scores formation

After extraction of the correlation matrix is completed, the factor loading matrix (source profiles) C can be formed as;

$$C = QE^{1/2} \tag{3.6}$$

Where $E^{1/2}$ is a diagonal eigenvalue matrix whose entries are the square roots of the eigenvalues. Only a first few important eigenvectors as indicated by the magnitude of associated eigenvalues are selected to form the C matrix.

Step 5. Varimax Rotation

The varimax mehod is the most commonly used method for orthogonal rotation of the factor axes. The goal of this method is to achieve a few loadings being of substantial size, while the rest is distributed randomly about zero for a given factor. The varimax rotation is performed by multiplying the target matrix with a transformation matrix (T_{vt}) or can be shown as

$$C_{vt} = CT_{vt} \tag{3.7}$$

where;

C	=	Unrotated factor loading matrix
C_{vt}	=	Varimax transformed factor loading matrix
T_{vt}	=	Varimax transformation matrix

In the case of two factors, the transformation matrix, T_{vt}, is of this form;

$$T_{vt} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{3.8}$$

where;

$$\tan 4\phi = \frac{2\left[m\sum\left(x^2 - y^2\right)(2xy) - \sum\left(x^2 - y^2 \sum(2xy)\right)\right]}{m\left[\sum\left[\left(x^2 - y^2\right)^2 - (2xy)^2\right]\right] - \left[\left[\sum\left(x^2 - y^2\right)\right]^2 - \left[\sum(2xy)\right]^2\right]} \tag{3.9}$$

Σ is a sum of all factor loadings (m) for the factors being rotated. The x and y are the value of factor loadings in sources (factors) 1 and 2, which are rotated pairwise.

3.5 Analysis of Data by Time Series Decomposition Technique

Decomposition analysis was applied only to PM_{10} since concentration in mg/m^3 of PM_{10} was often above TNAAQS (limit = 120 mg/m^3/24hr). Thus, we focus on the PM_{10} concentration characteristics with time.

This study applied decomposition method based on the assumption that the ambient PM_{10} concentration level was the product of trend, various cyclic, and irregular components. The model can be expressed as:

$$PM_{10} = T \times I \times C_1 \times C_2 \times C_3 \tag{3.10}$$

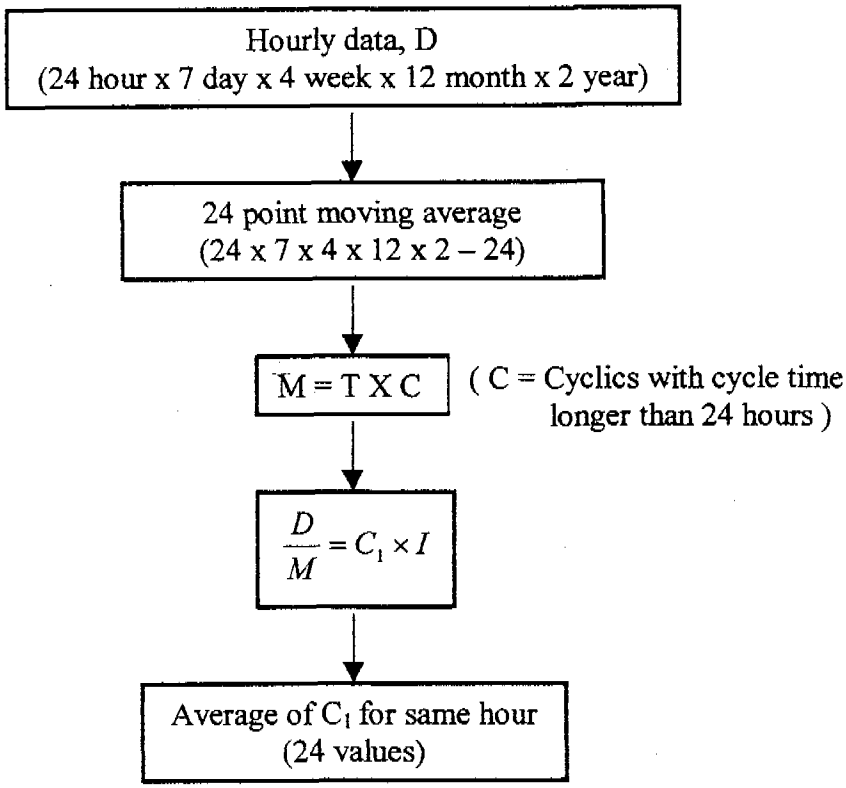
where;
T = the long-term trend
I = irregular component
C_1 = hourly cyclic component
C_2 = daily cyclic component
C_3 = monthly cyclic or seasonal component

Actual steps in the decomposition analysis are followed by decomposition method as described in the literature review section to obtain cyclic indices in addition to seasonal indices. However, steps in analysis of each cyclic are different in some parts such as input data and number of points used in moving averages.

3.5.1 Methodology for various cyclic indices computation

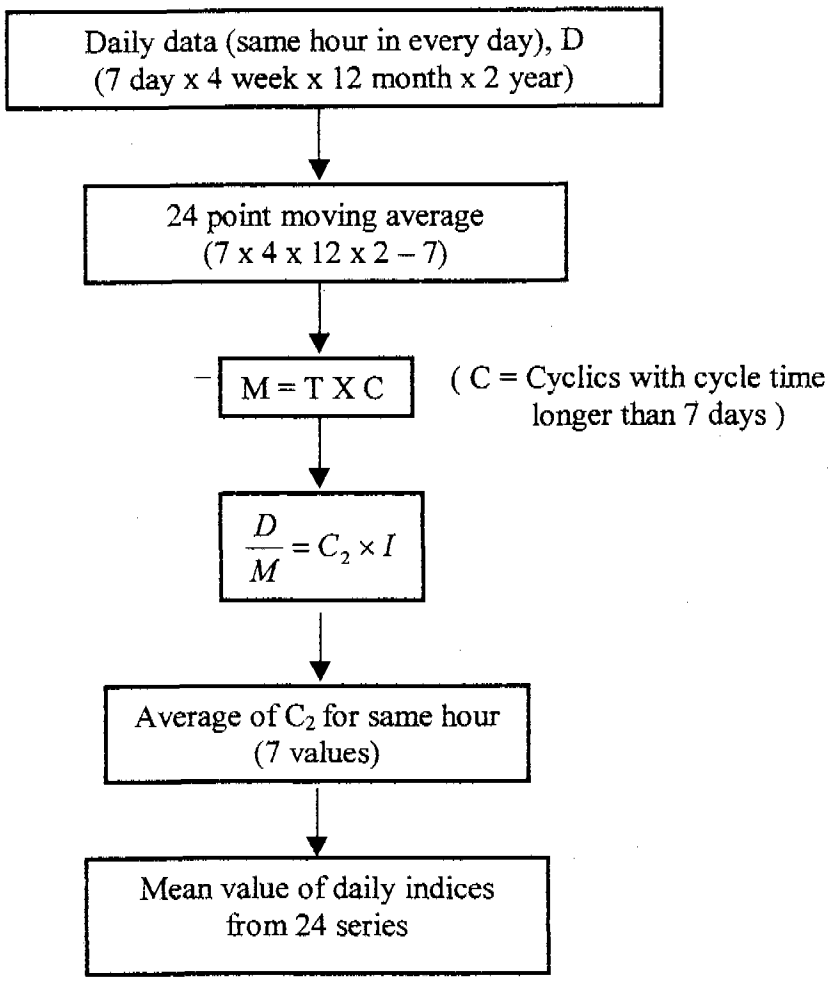
Cyclic$_1$ (hours/day)

Input data of cyclic$_1$ is hourly data, which has approximate 17,280 $(24\times7\times4\times12\times2)$ points. Cyclic$_1$ is computed based on the assumption that PM_{10} concentration level would occur in repetitious manner at every 24 hour so simple centered 24 point moving average was used. We isolate the cyclical behavior by dividing the original data by the 24 point moving average data. The result will be the cyclic$_1$ and irregular components. To eliminate the irregular component, Approximately 730 data of same hour in each day will be averaged to calculate hourly indices. Details on this analysis are as follows;



Cyclic$_2$ (days/week)

Cyclic$_2$ that is the daily cyclic index and is based on an assumption that PM_{10} concentration at same hour in each day was impacted from similar sources and cyclic$_2$ will repeat itself at every 7 days. Approximately 730 points are used for this cyclic and same hour in each day is chosen to represent daily data. Cyclic$_2$ will be run on 24 series of time such as series of 7:00, series of 8:00 etc. to correspond with our assumption. Simple centered 7 point moving average was used for cyclic$_2$ and divided original data by simple centered 7 point moving average to isolate cyclic$_2$ and irregular component. To calculate daily indices, we averaged cyclic$_2$ and irregular component in the same day and 7 values of daily indices for each specific hourly period will emerge. Finally, 24 series of daily indices will be averaged in order to represent the mean value of daily indices. Details on this analysis are as follows;



Cyclic$_3$ (months/year)

Input data of cyclic$_3$ is monthly data, which has 24 (12×2) points. Same hour in same month will be averaged over 30 data points to represent monthly data so cyclic$_3$ will be run on 24 series of monthly data. Cyclic$_3$ is computed based on the assumption that PM_{10} concentration level would occur in repetitious manner at every year so simple centered 12 point moving average was used in cyclic$_3$. We isolate the cyclical behavior by dividing the original data by the 12 point moving average data. The result will be the cyclic$_3$ and irregular components. As a result of insufficient monthly data, we cannot isolate irregular component so monthly indices may still contain irregularly. Finally, 24 series of monthly indices will be averaged in order to represent the mean value of monthly indices. Details on this analysis are as follows;