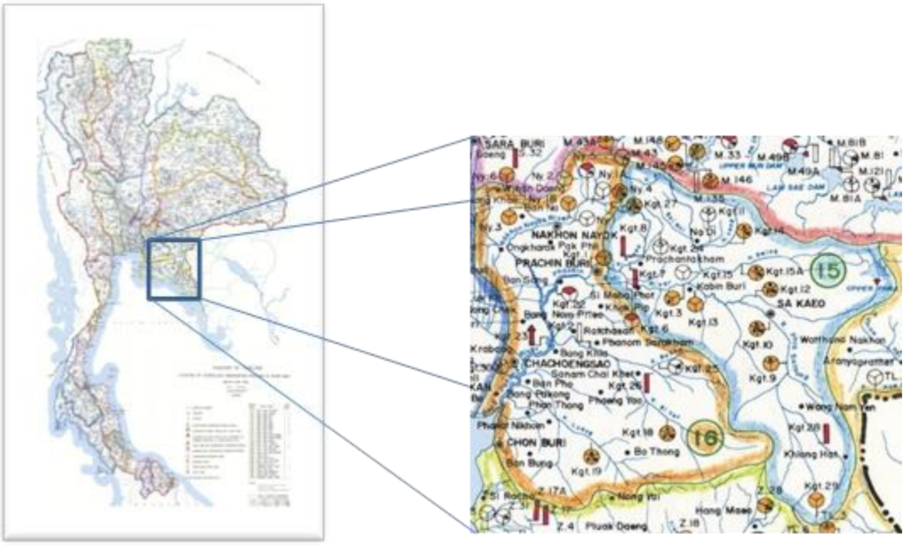
CHAPTER 2 THEORIES

2.1 Study Area

The study area is the Bang Pakong river basin**,** located at 60 kilometer to the east of Bangkok and Bang Pakong river basin, is one of the 25 river basins in Thailand. The details of physical and hydrological characteristics in the Bang Pakong river basin are as follows:

Topography : Bang Pakong river basin is one of the 25 river basins in Thailand and located adjacent to the east of the Chao Phraya river basin. The western boundary of the basin is located at the distance of about 60 kilometer east of Bangkok, and its catchment area is 17,660 square kilometer.



(Royal Irrigation Department of Thailand)

Figure 2.1 Location of study area**.**

2.2 Runoff

This study use monthly runoff data from the Hydrology Division of the Royal Irrigation Department of Thailand. Runoff is rain water falling on the ground and then flows along the surface down to the river after partial evaporation and seepage into the soil. Water flow along the surface is called overland flow, when it flows into the river, then it is called the Stream Flow. The usual amount of water flowing down the river is about 15% to 35% of the rainfall measured and depends on climate, soil water characteristics of the watershed areas and forests in the watershed area, etc..

Table 2.1 Runoff Station in Each Sub-basin of Bang Pakong River Basin

No.	Station Code	Station Name	Latitude	Longitude	D.A. (km^2)	Number of elements (N)	Missing data (%)	Record Length	
								Begin date	End date
1	KGT1	Prachin Buri	$14^\circ 03'01''N$	$101^\circ 22'03''E$	9,209	132	25	Apr 1966	Mar 1997
2	KGT3	Ban Ka bin Buri	$13^\circ 59'05''N$	$101^\circ 42'32''E$	7,425	864	25	Apr 1941	Mar 2003
3	KGT6	Si Maha Phot	$13^\circ 58'21''N$	$101^\circ 30'57''E$	7,978	36	-	Apr 1978	Mar 1981
4	KGT9	Ban Khao Chakan	$13^\circ 40'10''N$	$102^\circ 04'35''E$	2,264	492	7	Apr 1969	Mar 2013
5	KGT10	Ban Wang Khian	$13^\circ 48'29''N$	$102^\circ 03'35''E$	2,482	456	4	Apr 1966	Mar 2005
6	KGT12	Ban Kaeng	$13^\circ 56'02''N$	$101^\circ 58'41''E$	1,478	564	4	Apr 1966	Mar 2013
7	KGT13	Ban Nang Leng	-	-	5,347	348	5	Apr 1967	Mar 1997
8	KGT14	Ban Thung Faek	$14^\circ 09'30''N$	$101^\circ 52'52''E$	354	552	7	Apr 1966	Mar 2013
9	KGT15	Ban Rong Luai Khok Udom	$14^\circ 02'37''N$	$101^\circ 47'30''E$	789	108	23	Apr 1966	Mar 1975
10	KGT15A	Ban Kaeng Din So	$14^\circ 03'46''N$	$101^\circ 55'39''E$	548	528	9	Apr 1968	Mar 2013
11	KGT18	Ban Tha Kloi	$13^\circ 28'29''N$	$101^\circ 37'44''E$	1,078	420	5	Apr 1969	Mar 2004
12	KGT19	Ban Tha Bun Mee	$13^\circ 23'17''N$	$101^\circ 20'40''E$	473	480	4	Apr 1965	Mar 2006

Table 2.2 Runoff Station in Each Sub-basin of Bang Pakong River Basin (Cont.)

No.	Station Code	Station Name	Latitude	Longitude	D.A. (km^2)	Number of elements (N)	Missing data (%)	Record Length	
								Begin date	End date
13	KGT25	Ban Cham Pa Ngam	$13^{\circ}\,41'09''N$	$101^{\circ}\,36'32''E$	243	156	31	Apr 1959	Mar 1990
14	KGT27	Ban Khlong Yang	$14^{\circ}\,12'02''N$	$101^{\circ}\,22'05''E$	45	192	21	Apr 1983	Mar 1999
15	KGT33	Ban Sapan Hin	$14^{\circ}\,07'56''N$	$101^{\circ}\,43'52''E$	617	156	9	Apr 2000	Mar 2013
16	KGT42	Ban Ta Ra Pa	$13^{\circ}\,59'21''N$	$101^{\circ}\,57'30''E$	-	96	18	Apr 2005	Mar 2013

The monthly runoff data of Bang Pakong was collected from the Royal Irrigation Department of Thailand (RID). The availability and missing data of data each stations used in the study are shown in Table 2.4

Table 2.3 Length of time series and location where they were observed

No.	Station Code	Station Name	Number of elements	Number of Missing data	Begin Date	End Date
1	KGT1	Prachin Buri	132	33	Apr 1966	Mar 1997
2	KGT3	Ban Ka bin Buri	864	88	Apr 1941	Mar 2013
3	KGT6	Si Maha Phot	36	-	Apr 1978	Mar 1981
4	KGT9	Ban Khao Chakan	492	33	Apr 1969	Mar 2013
5	KGT10	Ban Wang Khian	456	19	Apr 1966	Mar 2005
6	KGT12	Ban Kaeng	564	21	Apr 1966	Mar 2013
7	KGT13	Ban Nang Leng	348	17	Apr 1967	Mar 1997
8	KGT14	Ban Thung Faek	552	41	Apr 1966	Mar 2013
9	KGT15	Ban Rong Luai Khok Udom	108	25	Apr 1966	Mar 1975
10	KGT15A	Ban Kaeng Din So	528	47	Apr 1968	Mar 2013
11	KGT18	Ban Tha Kloi	420	21	Apr 1969	Mar 2004
12	KGT19	Ban Tha Bun Mee	480	20	Apr 1965	Mar 2006
13	KGT25	Ban Cham Pa Ngam	156	48	Apr 1959	Mar 1990
14	KGT27	Ban Khlong Yang	192	41	Apr 1983	Mar 1999
15	KGT33	Ban Sapan Hin	156	14	Apr 2000	Mar 2013
16	KGT42	Ban Ta Ra Pa	96	17	Apr 2005	Mar 2013

Information of this monthly runoff is obtained from the Royal Irrigation Department of Thailand.

In Table 2.3 runoff considered the monthly runoff data of each station in Bang Pakong river basin, where missing data points used SSA for gab filling missing data. The idea of filling in missing data is to the idea of forecasting and consists in the continuation of the structure of the extracted component to the gaps caused by the missing data.

2.3 The Singular Spectrum Analysis (SSA)

The Singular Spectrum Analysis (SSA) is a data adaptive, nonparametric spectral estimation method based on embedding a time series $\{X_t = 1,\ldots,N\}$ in a vector space of dimension M .

2.3.1 Defragmenter

The process in this chapter is about the SSA method as follows:
Step 1: Embedding

$$X_t = [x_1 \quad x_2 \quad x_2 \quad \ldots \quad x_N]_{1\times N} \ , \tag{2.1}$$

$$X_t^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix}_{N\times 1} \ , \tag{2.2}$$

where: X_t is the matrix of the set data
t is time series, $t = 1, 2, 3, \ldots, N$

Format the matrix from matrix X_t^T that contains the data time series in the first column, a lag-1 shifted version of that time series in the 2^{nd} column, etc. in the following matrix:

$$P = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ x_2 & x_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ x_N & 0 & \cdots & 0 \end{bmatrix}_{N\times N} \tag{2.3}$$

where P is the matrix of lag time.

2.3.2 Covariance matrix

Step 2: Calculating $M \times M$ covariance matrix
The SSA method can create the covariance matrix C_X from matrix D as follows:

$$D = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ x_2 & x_3 & \cdots & x_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N'+1} & x_{N'} & \cdots & x_{N-1} \\ x_{N'} & x_{N'+1} & \cdots & x_N \end{bmatrix}_{N'\times M} \tag{2.4}$$

Transpose matrix D, written as follows:

$$D^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_{N'-1} & x_{N'} \\ x_2 & x_3 & \cdots & x_{N'} & x_{N'+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_M & x_{M+1} & \cdots & x_{N-1} & x_N \end{bmatrix}_{M\times N'} \tag{2.5}$$

where

$$C_X = \frac{1}{N'} D^T D \tag{2.6}$$

Substituting (2.4) and (2.5) in (2.6) we obtain from the equation (2.6) as follows:

$$C_X = \frac{1}{N'} \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ x_2 & x_3 & \cdots & x_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N'+1} & x_{N'} & \cdots & x_{N-1} \\ x_{N'} & x_{N'+1} & \cdots & x_N \end{bmatrix}_{N'\times M} \begin{bmatrix} x_1 & x_2 & \cdots & x_{N'-1} & x_{N'} \\ x_2 & x_3 & \cdots & x_{N'} & x_{N'+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_M & x_{M+1} & \cdots & x_{N-1} & x_N \end{bmatrix}_{M\times N'} \tag{2.7}$$

$$C_X = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_M \\ c_2 & c_3 & c_4 & \cdots & c_{M+1} \\ c_3 & c_4 & c_5 & \cdots & c_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_M & c_{M+1} & c_{M+2} & \cdots & c_{2M-1} \end{bmatrix}_{M\times M} \tag{2.8}$$

where: C_X is the $M \times M$ covariance matrix,
N is number of data,
M is number of columns by choosing from matrix,
N' is $N - M + 1$,
$N' = N - M + 1$ is number of each row in the matrix,
L is lag time of the matrix,
D is the matrix generated from sliding of a window size.

2.3.3 Eigenvalue and eigenvector

Compute eigenvalues and eigenvector of the covariance matrix C_X by decomposing eigenvalues, eigenvector and time coefficient from matrix C_X with dimension $M \times M$. Solving eigenvalues of covariance matrix C_X by finding λ satisfies the following:

$$C_X E_k = \lambda E_k \,,$$
$$(C_X E_k) - (\lambda E_k) = \bar{0} \,,$$
$$(C_X - \lambda I) E_k = \bar{0} \,,$$
$$(C_X - L) E_k = \bar{0} \,,$$

where: C_X is the covariance matrix,
E_k is the matrix of the eigenvector,
λ is the eigenvalues,
$\bar{0}$ is the zero vector,
I is the identity matrix.

Find the nonzero vector and scalar λ corresponding in the equation $C_X - L = 0$ can be obtained by:

$$C_X - L = 0,$$

$$\begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_M \\ c_2 & c_3 & c_4 & \cdots & c_{M+1} \\ c_3 & c_4 & c_5 & \cdots & c_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_M & c_{M+1} & c_{M+1} & \cdots & c_{2M-1} \end{bmatrix}_{M\times M} - \begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}_{M\times M} = \bar{0}, \quad (2.9)$$

$$\begin{bmatrix} c_1-\lambda & c_2 & c_3 & \cdots & c_M \\ c_2 & c_3-\lambda & c_4 & \cdots & c_{M+1} \\ c_3 & c_4 & c_5-\lambda & \cdots & c_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_M & c_{M+1} & c_{M+2} & \cdots & c_{2M-1}-\lambda \end{bmatrix}_{M\times M} = \bar{0}, \quad (2.10)$$

Since this is the $M \times M$ covariance matrix, we can use the formula given above to find its determinant.

$$|C_X - L| = 0$$

$$\begin{vmatrix} c_1-\lambda & c_2 & c_3 & \cdots & c_M \\ c_2 & c_3-\lambda & c_4 & \cdots & c_{M+1} \\ c_3 & c_4 & c_5-\lambda & \cdots & c_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_M & c_{M+1} & c_{M+2} & \cdots & c_{2M-1}-\lambda \end{vmatrix}_{M\times M} = 0 \quad (2.11)$$

Each eigenvalue $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_M$ corresponds to the eigenvector $E_1, E_2, E_3, \ldots, E_M$

$$\begin{matrix} \lambda_1 & \lambda_2 & \cdots & \lambda_M \\ \vdots & \vdots & \cdots & \vdots \\ E_1 & E_2 & \cdots & E_M \end{matrix}$$

where each vector $E_k = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix}$ has M component.

The vector E_k is the eigenvector of covariance matrix C_X, where $k = 1, 2, 3, \ldots, M$

The eigenvectors $E_k : k = 1, 2, 3, ..., M$ can be written in matrix form as:

$$E_k = \begin{bmatrix} e_1 & e_2 & e_3 & \cdots & e_M \\ e_2 & e_3 & e_4 & \cdots & e_{M+1} \\ e_3 & e_4 & e_5 & \cdots & e_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_M & e_{M+1} & e_{M+2} & \cdots & e_{2M-1} \end{bmatrix}_{M \times M} \tag{2.12}$$

where E_k is the matrix of the eigenvectors in which each column matrix is the eigenvector of matrix C_X. For the eigenvector, each mode is extracted from the original data.

2.3.4 Principal Components (PC)

To compute the principal components time series of each EOFs by defined matrix A, the matrix A is the principal components time series as follow in (2.13)

$$A = D_{N' \times M} E_{M \times N'} \tag{2.13}$$

This can be written in matrix form as

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_M \\ a_2 & a_3 & \cdots & a_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N'-1} & a_{N'} & \cdots & a_{N+1} \\ a_{N'} & a_{N'+1} & \cdots & a_N \end{bmatrix}_{N' \times M} \tag{2.14}$$

where A is the principal components time series (PC) for each column matrix A, is the principal components time series of each EOF.

2.3.5 Reconstruction (RC)

The process of reconstruction components (RC)
To compute the reconstruction a missing data, defined matrix R is the reconstruction as follow in

$$R = A_{N' \times M} E_{M \times M} \tag{2.15}$$

It can be written in matrix form as

$$R = \begin{bmatrix} a_1 & a_2 & \cdots & a_M \\ a_2 & a_3 & \cdots & a_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N'-1} & a_{N'} & \cdots & a_{N-1} \\ a_{N'} & a_{N'+1} & \cdots & a_N \end{bmatrix}_{N' \times M} \begin{bmatrix} e_1 & e_2 & e_3 & \cdots & e_M \\ e_2 & e_3 & e_4 & \cdots & e_{M+1} \\ e_3 & e_4 & e_5 & \cdots & e_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_M & e_{M+1} & e_{M+2} & \cdots & e_{2M-1} \end{bmatrix}_{M \times M} \tag{2.16}$$

By use linear combinations of these principal components and EOFs, They provide the reconstructed components (RCs) and the principal component time series
The reconstructed matrix R from the equation (2.15) can be written in matrix form as:

$$R = \begin{bmatrix} r_1 & r_2 & \cdots & r_M \\ r_2 & r_3 & \cdots & r_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N'-1} & r_{N'} & \cdots & r_{N-1} \\ r_{N'} & r_{N'+1} & \cdots & r_N \end{bmatrix}_{N' \times M} \tag{2.17}$$

Find average of each diagonal Matrix R and construct matrix R as follows:

$$RC = \begin{bmatrix} r_1 & r_2 & \cdots & r_M \\ r_2 & r_3 & \cdots & r_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N'-1} & r_{N'} & \cdots & r_{N-1} \\ r_{N'} & r_{N'+1} & \cdots & r_N \end{bmatrix}_{N' \times M} \tag{2.18}$$

2.4 Willmott's index of agreement

Willmott's index of agreement is a theory to determine error values obtained by model and values obtained from real surveys and is between -1 to 1, the values closer to 1 indicate that results obtained from the model are reliable but if the value is closer to -1 indicates that results obtained from the model are not reliable. (Willmott, Robenson and Matsuura, 2012) Willmott's index of agreement is found find as follows:

$$d_r = \begin{cases} 1 - \dfrac{\sum_{t=1}^{N} \left| P_t - O_t \right|}{2\sum_{t=1}^{N} \left| O_t - \overline{O} \right|}, & where \sum_{t=1}^{N} \left| P_t - O_t \right| \le 2\sum_{t=1}^{N} \left| O_t - \overline{O} \right| \\ \dfrac{2\sum_{t=1}^{N} \left| O_t - \overline{O} \right|}{\sum_{t=1}^{N} \left| P_t - O_t \right|} - 1, & where \sum_{t=1}^{N} \left| P_t - O_t \right| > 2\sum_{t=1}^{N} \left| O_t - \overline{O} \right| \end{cases}$$

where: d_r is the value obtained from Willmott theory (the value between -1 to 1),
N is the number data,
P_t is data prediction from the model at any time t,
O_t is data from real survey at any time t,
$\overline{O}$ is average of data from real survey.