CHAPTER 1 INTRODUCTION

1.1 Background

The Bang Pakong river basin an important rivers basin for in Thailand and has one of the 25 river basins in Thailand which it is located adjacent to the east of Chao Phraya river basin. Bang Pakong river basin is located at latitude $13^{\circ}\ 09^{'}\ N$ and longitude $14^{\circ}\ 32^{'}\ E$ with the western boundary of the basin located 60 kilometer East of Bangkok, and the catchment area about 10,707.48 square kilometer. The Bang Pakong river basin has annual runoff of about 3736 (unit is million cubic meters). Since each of the stations in the Bang Pakong river basin are asynchronous. The runoff is important for multi disciplines and education project, with the observed data of hydrological data being able to solve by some techniques. Salgado and Alonso (2006) proposed the Singular Spectrum Analysis (SSA) as a nonparametric technique of time series analysis based on principles of multivariate statistic. This technique decomposes a given time series into a set of independent additive time series. This method projects has original time series itself, following the procedure of principal component analysis (PCA). Missing data is a common problem in hydrological research, for spectral estimation. This research applied solved filling gaps of monthly runoff data of the Bang Pakong river basin during 1968 to 1992 using the Singular Spectrum Analysis (SSA) and choosing Ban Kaeng Din So station as a case study, since Ban Kaeng Din So station that has complete information. In performance test use two different experiments of filled in data by using SSA for Ban Kaeng Din So station, which the first experiment 1 is random data and experiment 2 is random cutting interval data with 10 cases being 4%, 5%, 7% ,9% ,10% 18%, 21%, 23%, 25% and 31%. Then data is filled-in by using SSA, with the results tested for efficiency by using Willmott's index of agreement.

Missing data is a common problem in hydrological research, which if solved can give the missing data can filled-in the gaps, and greater accuracy and better significance testing in the spectral analysis. This research applied gap-filling based on principles of multivariate statistics to fill in missing data. The missing data are iteratively estimated, thus removing the need for a prior assumption about the spatial form and parameters of the covariance matrix. Cross validation is used to determine the optimum number of leading EOFs to be retained for filling.

The Singular Spectrum Analysis (SSA) is a nonparametric technique of time series analysis based on principles of multivariate statistics. It decomposes a given time series into a set of independent additive time series. Fundamentally filled-in data by the method projects the original time series onto a vector basis obtained from the series, following the procedure of principal component analysis (PCA).

1.2 Literature Review

Kondrashov, Feliks, and Ghil. (2005) applied advanced spectra methods, Singular-Spectrum Analysis (SSA) and the Multi-Taper Method (MTM) to fill the gaps and to locate interannual and interdecadal periodicities. The gap filling uses a novel, iterative version of SSA. Our analysis reveals several statistically significant features of the records: a nonlinear, data-adaptive trend that includes a 256-year cycle, a quasi-quadrennial (4.2-year) and a quasi-biennial (2.2-year) mode, as well as additional periodicities of 64, 19, 12 and most strikingly 7 years.

Kondrashov and Ghil.(2006) used the Singular Spectrum Analysis (SSA) to fill the gaps in several types of data sets. For a univariate record, their procedure only uses temporal correlations in the data to fill in the missing points. For a multivariate record, multi-channel SSA (M-SSA) takes advantage of both spatial and temporal correlations, and iteratively produce estimates of missing data points, which are then used to compute a self-consistent lag-covariance matrix; cross-validation allows us to optimize the window width and number of dominant SSA or M-SSA modes to fill the gaps. The optimal parameters of our procedure depend on the distribution in time (and space) of the missing data, as well as on the variance distribution between oscillatory modes and noise.

Salgado and Alonso (2006) studied the applicability of SSA to signal processing for TCMS development. The signals used are those of vibration of the tool in two direction, and the wear parameter measured is the flank were (VB) because of its influence on the final dimensions and surface quality of the work piece. The estimation of the tool were performed by a multilayer neural network trained with the features extracted from the SSA-processed vibration signals.

Colebrook (1978) applied a form of SSA to biological oceanography and noted the duality between principal component analysis (PCA) in the space and time domain. Broomhead and King (1986: BK hereafter) applied the ``method of delays" of dynamical systems theory to estimate the dimension of and to reconstruct the Lorenz attractor using singular-value decomposition (SVD) on the trajectory matrix formed by lagged copies of a single series obtained from the system.

Vautard and Ghil (1989) realized the formal similarity between classical lagged-covariance analysis and the method of delays. They exploited this similarity further by pointing out the pairs of SSA Eigen modes corresponding to nearly equal eigenvalues and associated with principal components that are nearly in phase quadrature can represent efficiently a nonlinear, enharmonic oscillation. This is due to the fact that a single pair of data-adaptive Eigen modes can capture the basic periodicity of a boxcar or seesaw-shaped oscillation, rather than necessitate many overtones that will appear in methods with fixed basis functions.

Hassani (2007). The Singular Spectrum Analysis (SSA) technique is a novel and powerful technique of time series analysis incorporating the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. The possible application areas of SSA are diverse: from mathematics and physics to economics and financial mathematics, from meteorology and oceanology to social science and market research. Any seemingly complex series with a potential structure could provide another example of a successful application of SSA (Golyandinaet al., 2001). The aim of SSA is to make a decomposition of the original series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structure less noise. SSA is a very useful tool which can be used for solving the following problems: 1) finding trends of different resolution; 2) smoothing; 3) extraction of seasonality components;4) simultaneous extraction of cycles with small and large periods; 5) extraction of periodicities with varying amplitudes; 6) simultaneous extraction of complex trends and periodicities; 7) finding structure in short time series; and 8) change-point detection solving, all these problems corresponds to the basic capabilities of SSA. To achieve the above mentioned capabilities of SSA, we do not need to know the parametric model of the considered time series.

Zhigljavsky and Golyandina (2013), Singular spectrum analysis (SSA) is a technique of time series analysis and forecasting. It combines elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. SSA aims at decomposing the original series into a sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and a structure less noise. It is based on the singular-value decomposition of a specie matrix constructed upon time series. Neither a parametric model nor stationary type conditions have to be assumed for the time series; this makes SSA a model-free technique.

Vautard et al. (1992), the window length M should be chosen to be longer than number of data points in the oscillatory periods under investigation, and shorter than number of data points in the spells of an intermittent oscillation. Recommend that the window length be less than about $\frac{N}{5}$ where N is the number of points in the time series. Robustness of results to M is an important test of their validity. The choice of window length sets the dimension of the lag autocorrelation matrix to be constructed and diagonalized by SSA, and thus determines the computational burden of the application.Larger values of M correspond to higher spectral resolution, although there is no direct equivalence between them.

1.3 Research Objective

The objectives of this research is to analyze the gap filling of monthly runoff data using Singular Spectrum Analysis (SSA) for a case study of Bang Pakong river basin.

1.4 Research Scopes

In this research, gap-filling of monthly runoff data of Bang Pakong basin problem will be analyzed using the Singular Spectrum Analysis (SSA), which choose the comparison between case random (experiment1) data and random cutting interval data (experiment 2) to choose the optimum of a window size and SSA component filled-in data.

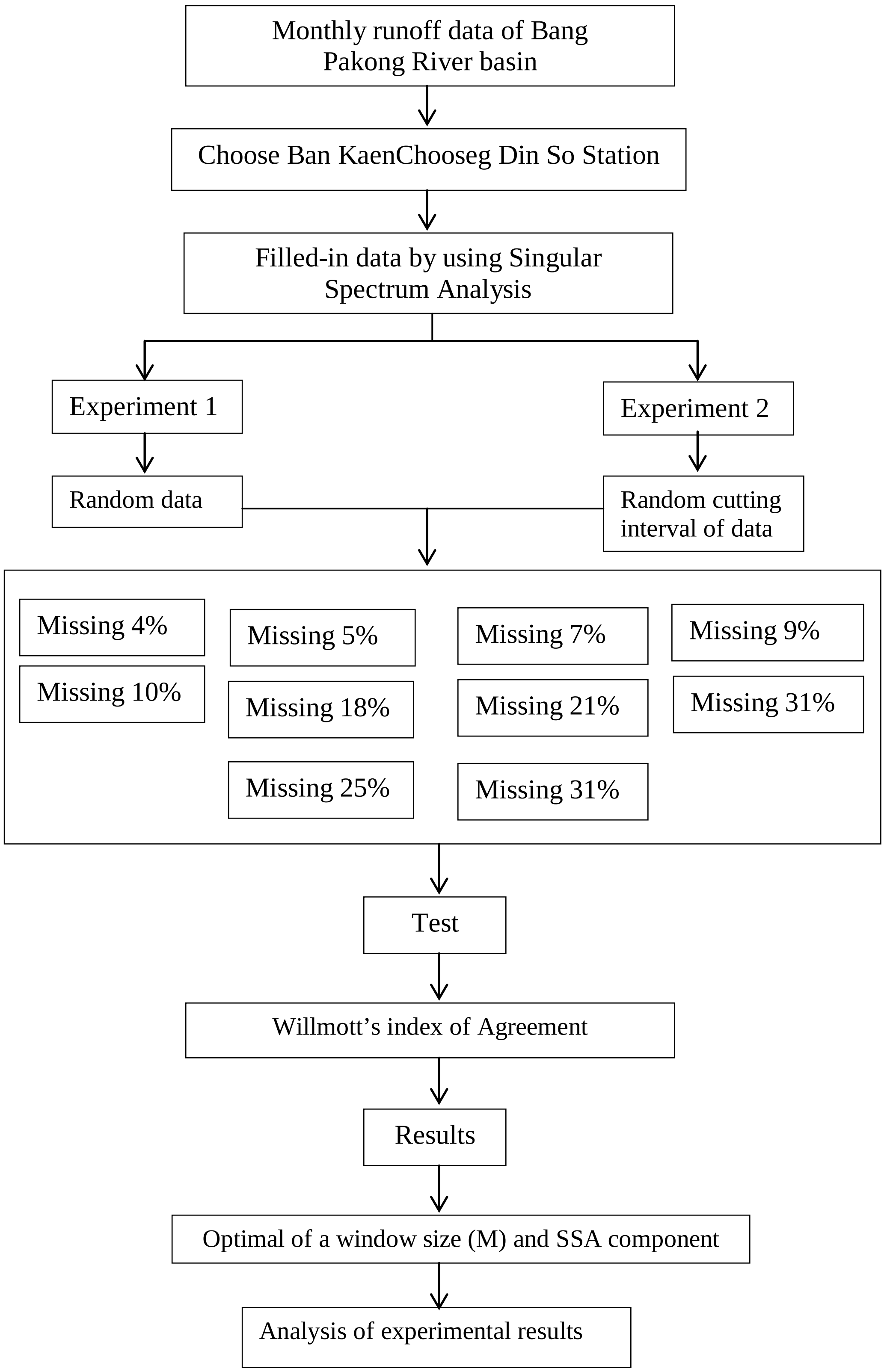

Figure 1. 1 Research flowchart