Wavelet application for improving annual rainfall prediction and investigation of monthly rainfall distribution over a year

S.Tantanee
Civil Engineering Department, Naresuan University

Abstract

The annual rainfall prediction is significant in planning for reservoir operation and irrigation areas. In recent years, the technique of wavelet has been widely applied to various water resources research categories because of its time-frequency representation. This study was undertaken to improve annual rainfall prediction of the conventional autoregressive (AR) model by applying wavelet transformation. The 52 year rainfall records of 4 stations distributed over the northeastern part of Thailand were analyzed by the proposed wavelet-AR model (WARM). Two rainfall variables, the number of rainy days per year and the annual rainfall were analyzed in this study. Comparing the obtained R square from WARM with conventional AR, it can be seen that WARM can improve the result of annual rainfall prediction. By applying wavelet to monthly rainfall series, the technique of Unit Disaggregation Curve (UDC) has been developed in this study in order to investigate the pattern of rainfall distribution within a year. The UDC is expressed in the unit curve that is able to represent that pattern of rainfall distribution graph for a year.

Keywords: wavelet, autoregressive model, rainfall, rainfall prediction

1. Introduction

For water resources planning purposes, a long-term series of rainfall is required in hydrological and simulation models. There have been many attempts to find the most appropriate method for rainfall prediction, for example: coupling physical, marine, and meteorological or satellite data with the forecasting model or even applying several techniques such as artificial neural network (ANN) or fuzzy logic as a forecasting approach. In recent years, several numerical weather forecasts have been proposed for weather prediction but most of these models are limited to short period forecasts. There have been several time series models proposed for modeling annual rainfall series, such as autoregressive model (AR) (Yevjevich,1972), fractional Guassian noise model (Matalas and Willis, 1971), autoregressive moving-average models (ARMA) (Carlson et. al.,1970) and disaggregation multivariate model (Valencia and Schaake,1973) etc. In the past decade, the theory of wavelet has been introduced to signal processing analysis. The wavelet transform has recently been successfully applied to wave data analysis and other oceanographic engineering applications (e.g. Massel, 2001; Teisseire et al., 2002; and Huang M.-C., 2004), such as the time-frequency character of long-term climatic data investigated using the Continuous Wavelet Transform (CWT) technique (Lau and Weng, 1995; Torrence and Compo, 1997; Mallat, 1998) and wavelet analysis of wind wave measurement obtained from a coastal observation tower during period of 2000-2001 (Huang ,2004) etc. The advantage of wavelet is that it provides a mathematical process to decompose the signal into multi-levels of details so that a multiresolution analysis can be carried out (Liang and Page, 1997). Wavelet transforms were also applied to time series prediction preprocessed for multi-step prediction (Fu-Chiang Tsui et al.,1997).

This study introduces an approach for improving annual rainfall prediction and proposes a technique for investigating the rainfall distribution pattern within a year. The wavelet decomposition process was applied to an annual rainfall series to separate the rainfall series into 2 parts: the high-scale, low-frequency components (the approximations, A) and the low-scale, high-frequency components (the details, D). The approximations which are the part that show the identity of a series can easily be stated as a trend whereas the details parts are a nuance of the rainfall series. Then, these two parts were analyzed separately as time series using an AR model from which predicted annual values can be obtained. With the reconstruction process, both predicted A and D are merged to construct the final predicted annual rainfall. The technique of unit disaggregating curve (UDC) was also developed in this study to investigate the rainfall distribution pattern over a year. As the approximations part demonstrates the identity of the rainfall series, this part of monthly rainfall records were analyzed to obtain rainfall distribution in each month within a year. The UDC is represented in a unit curve that can easily be applied to any level of annual rainfall to predict the monthly rainfall in that year.

2. Studied data

As the northeastern part of Thailand is the aridest area in the country and often faces the problem of water resources management, the effective process of rainfall prediction will be helpful for water resources planning in this area. Therefore, the rainfall of this region is selected to analyze in this study. The rainfall records over the area have been collected since 1951 by the Thai Meteorological Department (TMD) which provides long enough rainfall records to analyze. In order to study the prediction model that can be applied for rainfall over the northeastern region, the stations were random selected from the upper, middle and lower part and also from different watersheds and different characteristic feature of areas. The rainfall data from the 1951-2002 records of 4 gauging stations distributed over the northeastern part of Thailand were selected to analyze by the proposed wavelet-AR model (WARM). The locations and details of these stations are shown in Figure 1 and Table 1.

3. Method and model structure

3.1 Autoregressive model

The autoregressive (AR) model has been extensively applied to hydrology and water resources analysis for a long time. The AR is a model of time dependence, where the value of variables at the present time depends on the values at a previous time. AR models may have constant parameters, parameters varying with time or combination of both. For AR with constant parameters, a stationary time series y_t normally distributed with mean μ and variance σ^2, the AR of order p, denoted by AR (p), can be represented as:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + ... + \phi_p(y_{t-p} - \mu) + \varepsilon_t \qquad (1)$$

$$y_t = \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \varepsilon_t \qquad (2)$$

where y_t is time dependent series and ε_t is the time independent series which is uncorrelated with y_t. The series of y_t is also normally distributed with mean zero and variance σ_ε^2. The coefficients $\phi_1,\ldots,\phi_p$ are called autoregressive coefficients. Various forms of AR models, which have been used in the field of stochastic hydrology, represent the same autoregressive process. (Fiercing and Jackson,1971; Yevjevich, 1972; and Box and Jenkins,1970) The first order autoregressive (lag-one autoregressive) or first order Markov can be expressed as:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t \qquad (3)$$

In order to yield a normal series for annual time series, y_t, it is necessary to transform these variables before carrying out the statistical analysis. The transformation of y may be a simple logarithm transform, $y' = \log(y)$ or z transform, $z = y - \bar{y}$.

The consideration of p in this study was undertaken by examination of the correlation between the rainfall at t (y_t) and rainfall at t-1 (y_{t-1}), t-2 (y_{t-2}),.., t-p (y_{t-p}) . In this study, it is found that there are the correlation for annual rainfall between y_t and y_{t-p} up to p = 4. Then, the lag-four autoregressive was used in this study.

3.2 Wavelet transform for signal decomposition

Normally, it is difficult to get information from the raw signal; therefore the mathematical transformations are usually applied to a signal to get further information. The most well known transformation is Fourier transforms. Although, the windowed Fourier transform (WFT) is used to maintain time and frequency localization in signal analysis, it still has a problem in inconsistent treatment of different frequencies which cause the frequency localization to be lost at low frequencies and time localization is lost at high frequencies. By decomposing or transforming a one-dimensional time series into a diffuse two-dimensional time-frequency image simultaneously, wavelet analysis can solve this problem. It can get information on both the amplitude of any periodic signals within the series, and how this amplitude varies with time. In wavelet analysis, a mother wave $\Psi(t)$ and a linear combination of its dilated and shifted versions are used to represent a given signal (Burrus et al.,1998). The wavelet system is a two-dimensional expansion set (basic) for some class of one-dimensional signal. The wavelet expansion provides a time-frequency localization of the signal. First-generation wavelet systems are generated from a single scaling function (wavelet) by simple scaling and translation. The two dimensional parameterization is achieved from the function (mother wave) $\psi(t)$ by

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \quad \text{j,k} \in Z \qquad (4)$$

where Z is the set of all integers and the factor $2^{j/2}$ maintains a constant norm independence of scale j. The scale index j indicates the wavelet's width. This parameterization of the time or space location by k and the frequency or the logarithm of scale by j turns out to be effective.

This two variable set of basic functions is used in a way similar to the short-term Fourier transform. In order to generate a set of expansion functions such that any signal in the space of square integrable functions, ($L^2(\boldsymbol{R})$) can be represented by the series

$$f(t) = \sum_{j,k} a_{j,k} 2^{j/2} \psi(2^j t - k) \qquad (5)$$

Where the two dimensional set of coefficients $a_{j,k}$ is called the discrete wavelet transform (DWT) of *f(t)*. It is desired that wavelet basic functions be orthonormal in order to simplify the computation of the coefficients. The wavelet coefficients $a_{j,k}$ are obtained as

$$a_{j,k} = \left\langle f(t), \psi_{j,k}(t) \right\rangle \qquad (6)$$

If the $\Psi_{j,k}(t)$ form an orthonormal basis for the space of signals of interest, the discrete wavelet transform can be expressed as

$$a_{j,k} = \sum_{t} f(t) \psi_{j,k}(t) \qquad (7)$$

From (4), (5) and (7) the set of expansion function can be represented by

$$f(t) = \sum_{j,k} \left\langle \psi_{j,k}(t), f(t) \right\rangle \psi_{j,k}(t) \qquad (8)$$

where *f(t)* is the signal to be analyzed, $\psi_{j,k}(t)$ is the dilated and shifted version or the basis expansion functions of mother wave $\Psi(t)$, j and k are integer indices.

For DWT, the wavelet coefficients, which are considered as filters, are placed in a transformation matrix. This matrix is applied to the raw data vector (Graps, 1995). The coefficients are ordered using two dominant patterns, one that works as a smoothing filter (like moving average), and one pattern that works to bring out the data’s detailed information. The matrix is efficiently applied to the data vector by a hierarchical algorithm, sometimes called a pyramidal algorithm (Mallat, 1989). The output of DWT consists of the remaining smooth component and the accumulated detailed component. The decomposition process can be iterated, so that one signal is broken down into many lower resolution components. The wavelet decomposition tree is shown in Figure 2(a). Since the analysis process is iterative, in theory it can be continued indefinitely. In practice, the suitable number of levels is based on the nature of the signal. These 2 components can be assembled back into the original signal without loss of information. This process is called reconstruction, or synthesis. The mathematical manipulation that affects synthesis is called the inverse discrete wavelet transform (IDWT). Figure 2(b) shows the reconstruction process.

3.3 Model structure for rainfall prediction

The concept of the proposed model is based on coupling AR with wavelet transforms (wavelet-AR model, WARM). In this study, there are 2 parts of wavelet involved; decomposition and reconstruction process. The principle of the model is to split the signal into high frequency and low frequency components. The wavelet transform was applied to a rainfall series in order to decompose the original signal (annual rainfall series) into the low frequency, approximations (A,) and the high frequency, details (D). The approximations, which are the smoothed signal, represent the trend of the rainfall series whereas the details represent the nuance of the rainfall series. The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. The wavelet decomposition can yield valuable information about the signal.

The process of 3-level decomposition with wavelet is demonstrated in Figure 3 as a wavelet decomposition tree. After the annual rainfall series was separated into 2 series of approximations and details, the predicted A and D were analyzed by the AR model. Then these 2 predicted A and D were reconstructed with IDWT and the predicted annual rainfall was obtained. The schematic diagram of WARM is shown in Figure 4. A comparison of the predicted result among WARM, AR, AR with z-transform and AR with a logarithm transform can be done by using the R-square fitting function which is the square of the Pearson product moment correlation.

$$R^2 = \frac{(n\sum_{j=1}^{n}(y_j\hat{y}_j) - (\sum_{j=1}^{n}y_j)(\sum_{j=1}^{n}\hat{y}_j))^2}{[n\sum_{j=1}^{n}y_j^2 - (\sum_{j=1}^{n}y_j)^2][n\sum_{j=1}^{n}\hat{y}_j^2 - (\sum_{j=1}^{n}\hat{y})^2]} \quad (9)$$

where y_j is the observed data at j , $\hat{y}_j$ is the predicted value at j and n is the number of data.

3.4 Investigation of rainfall distribution over a year with a unit disaggregation curve (UDC)

As the approximations are the smoothed signal and demonstrate the gross view of a signal, the investigation of rainfall distribution over a year should be considered from this component. The wavelet transform was applied to the rainfall series of each month in order to decompose into 2 components of the details and the approximations. The obtained approximations component of each month was plotted over a year to demonstrate characteristics of rainfall distribution within a year. There are 2 methods to get the distribution curve of monthly rainfall: (1) by fitting these plotted data with a polynomial curve fitting and; (2) by constructing an average value curve of plotted data. The UDC technique was developed from the basic concept of the linear relationship between monthly rainfall and annual rainfall. The relationship of monthly and annual rainfall can be expressed as:

$$y = \sum_{n=1}^{12} x_n \quad (10)$$

where y is the annual rainfall and x_n is the monthly rainfall at month n. The obtained distribution curve was rearranged into the accumulation rainfall curve which can express the accumulative rainfall of month n. Then, the unit disaggregation curve (UDC) is constructed by altering this accumulative rainfall curve to a unit curve that is easily done by dividing it by the annual rainfall.

The process of UDC construction is shown in Figure 5.

4. Results and discussion

The sequences of the annual rainfall (mm) from 4 selected stations have been analyzed with WARM in this study. The series generated with WARM were compared with those from the conventional AR model, AR with log-transforms and AR with z-transforms. Figure 6 represents the predicted results of each model compared with the historical series. It is apparent that WARM with 1-3 levels of filter can represent the characteristics of the annual rainfall series, whereas three models of AR can explain only the trend of predicted rainfall compared with the previous values. The extreme values can not be detected by using AR or AR with transforms. By comparing obtained R-square (R^2) in table 2, it is obvious that all of WARMs with a filter provide better result in prediction than the traditional AR and AR with transforms. Among the models tested, the WARM with 3 levels of filter is the most appropriate model for annual rainfall prediction.

The proposed UDC technique is to investigate rainfall distribution over a year. The monthly rainfall in n^{th} months for a 52 year record were applied with wavelet transforms in order to get the approximations values. Figure 7 shows both of the plotted approximations(A) and details (D) of monthly rainfall at n^{th} months over the year of records. The distribution curve was then constructed from these plotted data by (1) polynomial fitting and (2) average value of data in each month as the example also shows in Figure 7. By visual considering the characteristics of the obtained A distribution curve, it is can be seen that after applying 3 levels of decomposition the trend of the A-curve disappears. It can be concluded that the most appropriate level of decomposition should be only 2. Therefore, the distribution curve of 2 levels of decompositions was utilized to construct the unit disaggregation curve (UDC). The obtained UDC is shown in Figure 8 for monthly rainfall distribution over a year. It can be seen from the figures that there are slight differences between the UDC obtained from a polynomial fitting curve and the average value curve. As UDC is analyzed from the approximations component of the rainfall series, this distribution curve can describe the gross view of rainfall that is distributed over a year. It should be noted that if UDC is applied to the monthly rainfall prediction, it is necessary to consider in detail the details component or nuance of rainfall.

5. Conclusions

WARM is the alternative for improving the AR model by applying wavelet transforms. The process can be used the same as the application of traditional AR to predict the annual rainfall for a year for the purpose of long-term water resources planning. According to the need of rainfall prediction in reservoir operation simulation models, WARM can be applied to predict a long range rainfall series to serve those models in order to make the irrigation or water resources planning management more effective**.**

Studying 4 selected rainfall stations data records, the obtained UDC from analyzing monthly rainfall can illustrate the pattern of rainfall distribution within a year. The UDC is expressed in a unit curve that can be applied to any level of annual rainfall for downscaling to the monthly rainfall. It can be concluded that the UDC technique is the first step of downscaling annual rainfall to monthly rainfall prediction. However, this UDC can only be used to explain the gross view or trend of rainfall. Further study should be undertaken to consider more details on the nuance of rainfall series which is unavoidable in the rainfall prediction model.

References

Box G.E, Jenkin G. 1970. Time series analysis, forecasting and control. Holden-Day Inc., San Francisco.

Burrus C.S, Gopinath R.A, Gao H. 1998. Introduction to wavelet and wavelet transform. Prentice-Hall International, Inc.

Carlson R.F, MacCormick A.J.A, Watts D.G. 1970. Application of linear models to four annual streamflow series. Jour. Water Resour. Res., 6,4: 1070-1078.

Daubechies I. 1993. Orthomormal bases of compactly supported wavelet II, variations on a theme. SIAM Journal of mathematical analysis, 24(2): 499-519.

Feircing M.B, Jackson B.B. 1971. Synthetic hydrology monograph no.1. American geophysical Union, Washington D.C.

Graps A. 1995. An introduction to wavelets. IEEE computational science and engineering, vol.2,2.

Huang M.-C. 2004. Wave parameters and functions in wavelet analysis. Ocean Engineering 31 (1): 111-125.

Lau K.M, Weng H.Y. 1995. Climate signal detection using wavelet transform: how to make a time series sing. Bull. Amer. Meteor. Soc. 76: 2391-2402.

Liang Y, Page E.W. 1997. Multiresolution learning of signal processing. 45,11: 2858-2864

Mallat S. 1998. A wavelet tour of signal processing. San Diego, CA: Academic Press.

Matalas N.C, Willi J.R. 1971. Statistical properties of multivariate fractional noise process. Jour. Water Resour. Res., 7,6: 1460-1468.

Massel S.R. 2001. Wavelet analysis for processing of ocean surface wave records. Ocean Engineering 28 ; 957- 987.

Teisseire L.M, Delafoy M.G, Jordan D.A, Miksad R.W, Weggel D.C. 2002. Measurement of the instantaneous characteristics of natural response modes of a spar platform subjected to irregular wave loading. International Journal of Offshore and Polar Engineering 12(1):16-24.

Torrence C, Compo G.P. 1997. A practical guide to wavelet analysis. Bull. Amer. Meteor. Soc. 79(1): 61-78.

Tsui F.-C, Li C.-C, Sun M, Sclabassi R.J. 1997. A comparative study of two biorthogonal wavelet transforms in time series prediction. IEEE: 1791-1796.

Valencia D.R, Schaake J.C.Jr. 1973. Disaggregation processes in stochastic hydrology, Jour. Water Resour. Res. 9, 3: 580-585.

Yevjevich V. 1972. Stochastic processes in hydrology. Water Resources publication, Fort Collins, Colorado.

Table 1 Details of 4 selected rainfall stations

Station	Period of record	Latitude–Longitude	No. Year of record	Climatic Annual mean (mm)	Mean number of wet days
381201 Khon Kaen	1951-2002	16.26N 102.50E	52	1212.5	87.21
431201 Nakhon Ratchasima	1951-2002	14.58N 102.05E	52	1092.6	86.73
407501 Ubon Ratchathani	1951-2002	15.15N 104.52E	52	1580.8	100.88
354201 Udon Thani	1951-2002	17.23N 102.48E	52	1460.3	101.42

Table 2 Obtained R-square from models for annual rainfall

Model	Obtained R^2 of annual rainfall			
	Khon Kaen	Nakhon Ratchasima	Ubon Ratthathani	Udon Thani
WARM-1 level Filter	0.9228	0.8954	0.9210	0.8907
WARM- 2 level filter	0.9568	0.9346	0.9633	0.9346
WARM- 3 level filter	0.9603	0.9404	0.9684	0.9425
AR model	0.1632	0.0811	0.1263	0.1940
AR with z-transformed	0.0654	0.0854	0.1106	0.0416
AR with Log-normalized	0.1539	0.1048	0.1321	0.1925

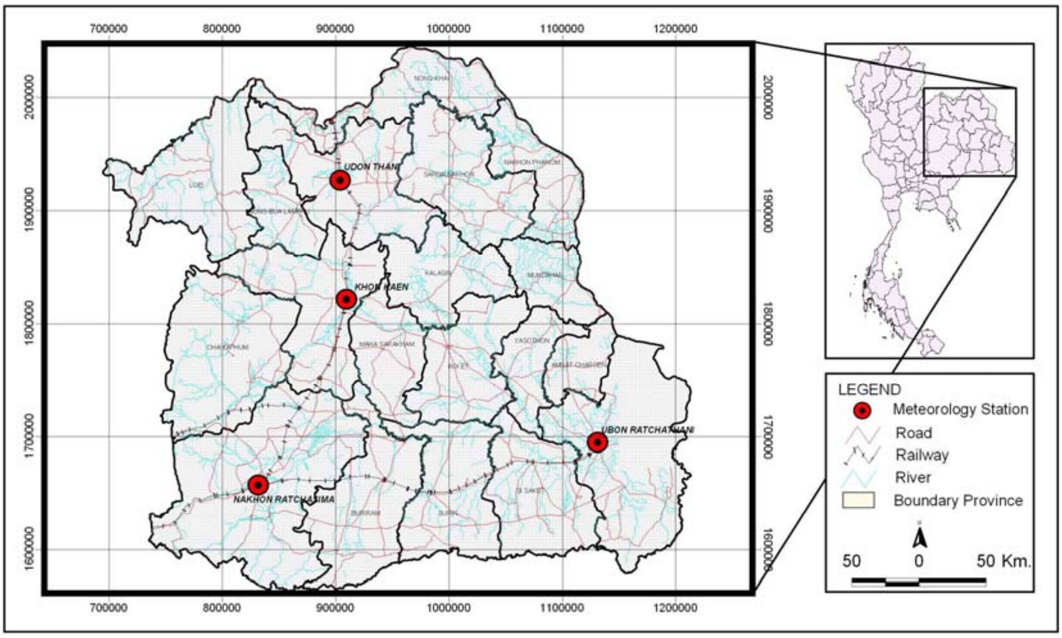
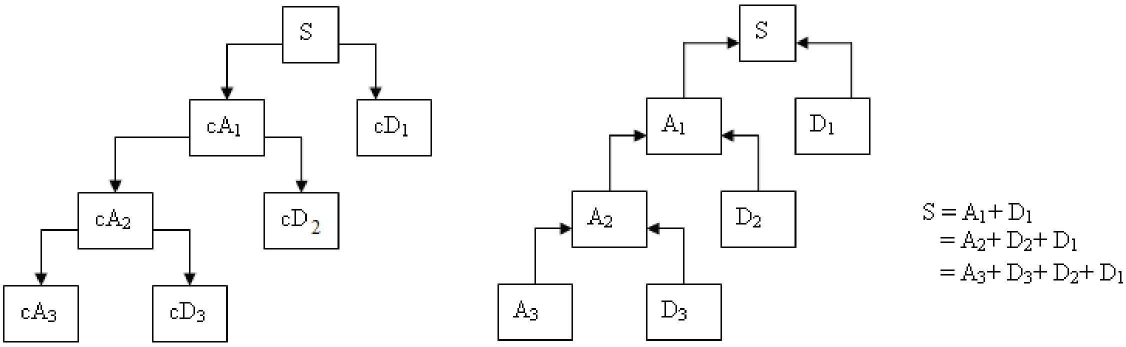


Figure 1 Location of the 4 selected rainfall stations



(a) Decomposition process

(b) Reconstruction process

Figure 2 Decomposition process and reconstruction process for wavelet analysis

S= initial signal; A_x = approximation component of x decomposition level;

D_x= details component of x decomposition level; c= wavelet coefficient

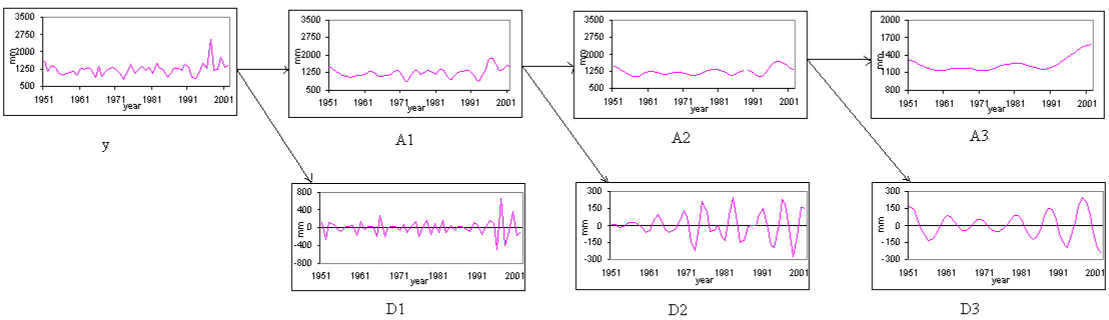


Figure 3 Wavelet decomposition tree of annual rainfall series

y= initial signal; A_x = approximation component of x decomposition level;

D_x= details component of x decomposition level

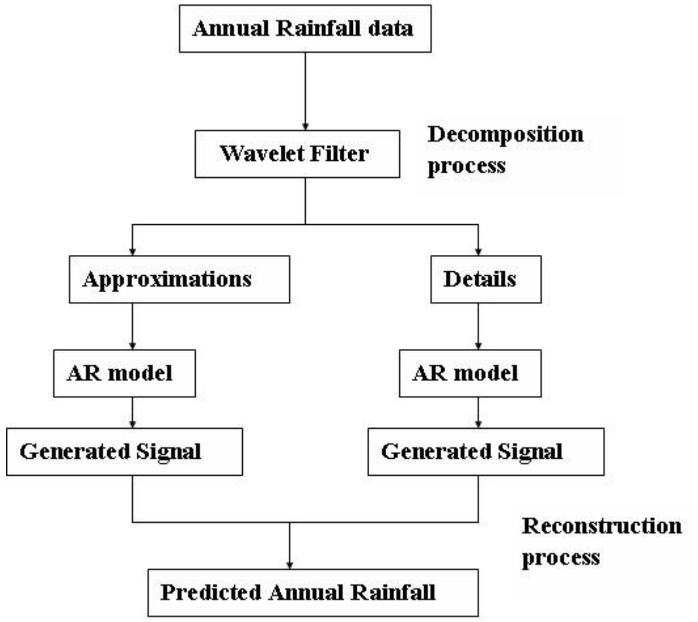


Figure 4 Methodology structure of WARM

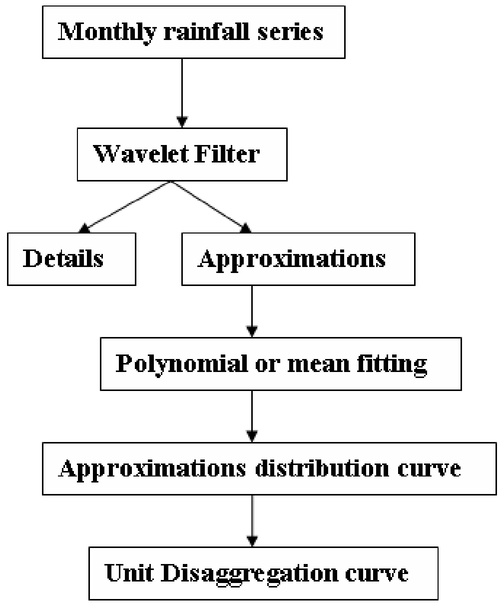


Figure 5 The UDC construction process

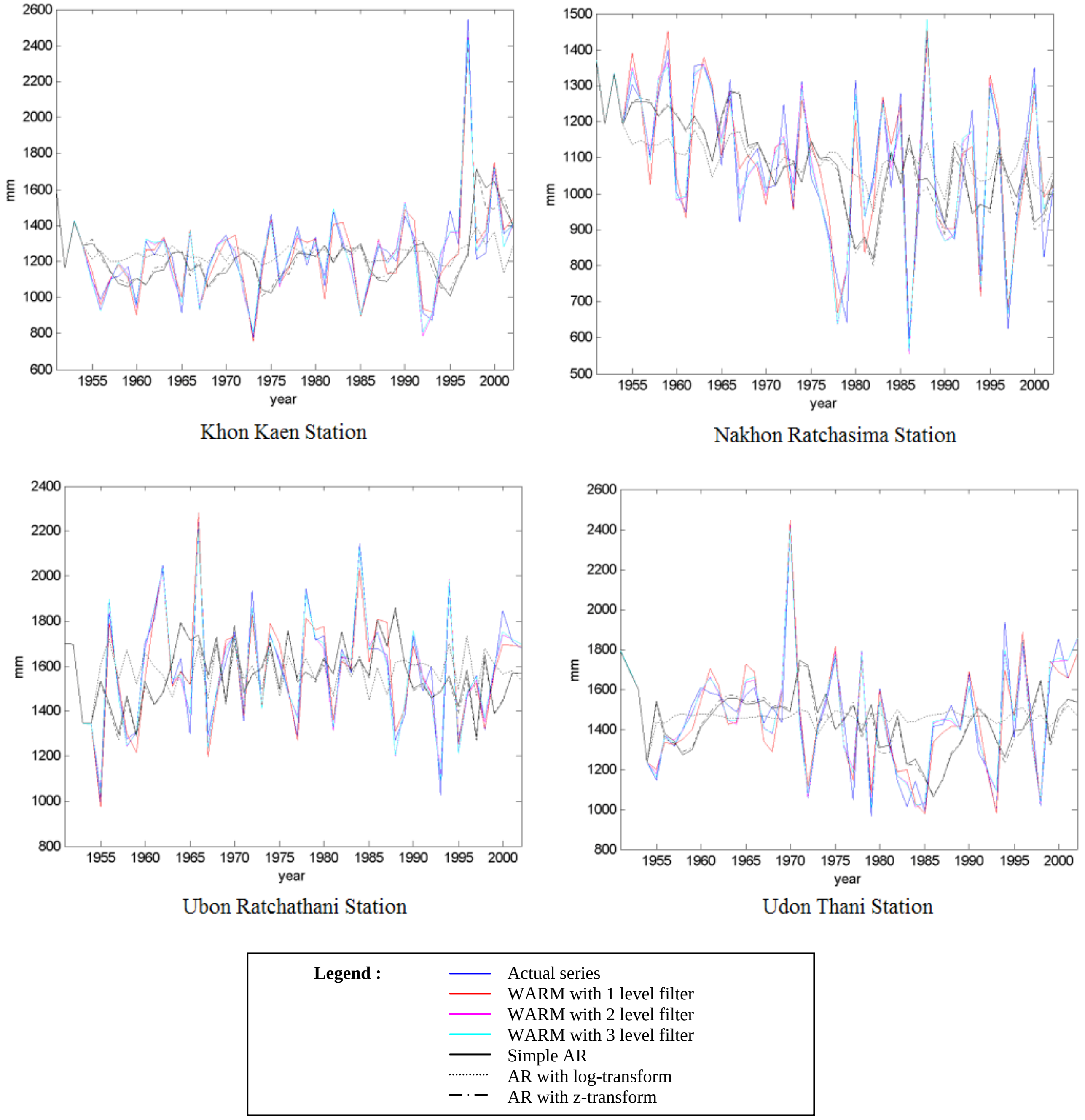


Figure 6 Comparison of the studied model with historical data for annual rainfall, mm

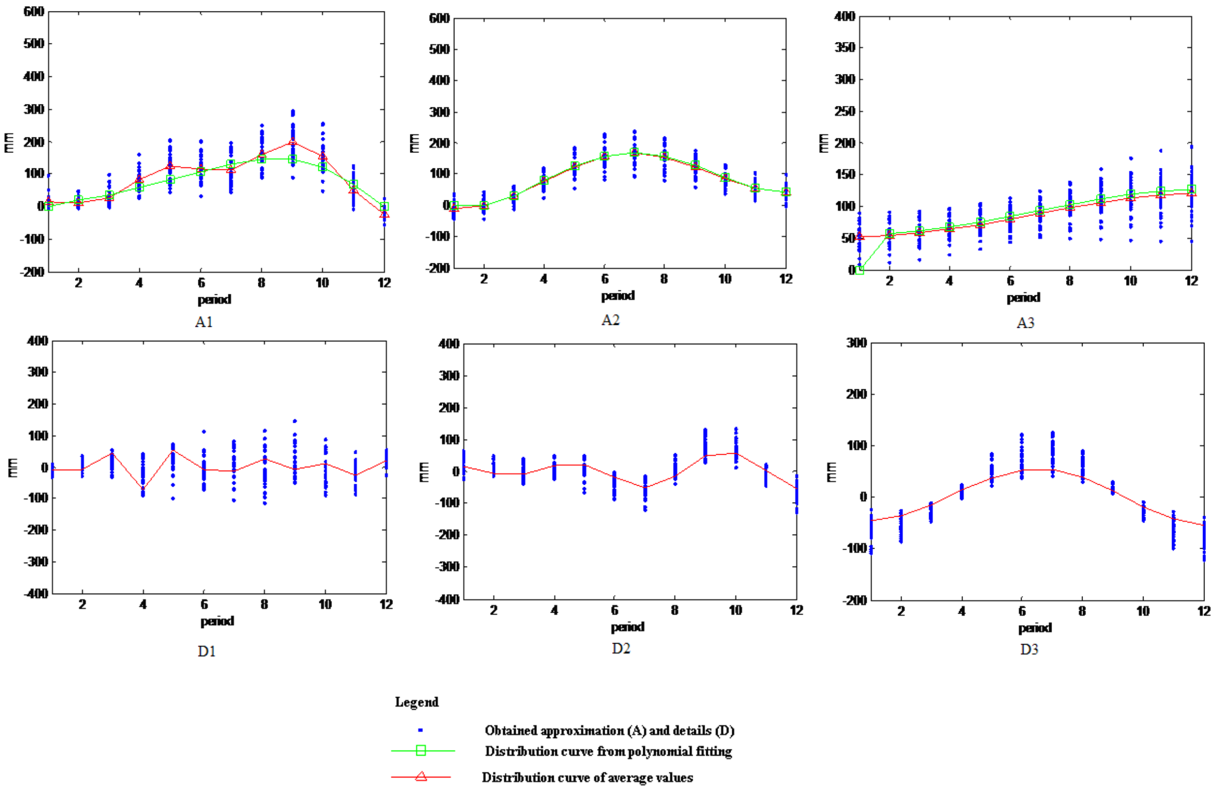


Figure 7 Obtained approximations (A) and details (D) of the monthly rainfall over a year at Nakhon Ratchasima station:

Ax = Obtained A with x level of decomposition

Dx = Obtained D with x level of decomposition

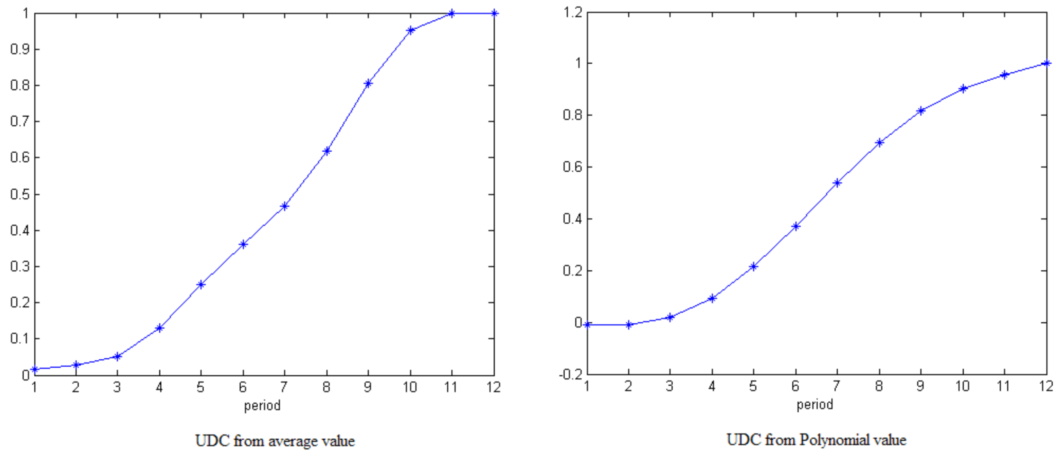


Figure 8 UDC of the monthly rainfall over a year obtained from 2 levels of wavelet decomposition (Nakhon Ratchasima station)

ADDRESS FOR CORRESPONDENCE
Sarintip Tantanee
Department of Civil Engineering
Naresuan University
Tha Pho, Muang, Phitsanulok 65000
THAILAND
E-mail: stantanee@yahoo.com, sarintipt@nu.ac.th