JOURNAL OF
ENVIRONMENTAL HYDROLOGY

The Electronic Journal of the International Association for Environmental Hydrology
On the World Wide Web at http://www.hydroweb.com

VOLUME 13 2005

COUPLED WAVELET-AUTOREGRESSIVE MODEL FOR ANNUAL RAINFALL PREDICTION

S. Tantanee[1]
S. Patamatammakul[1]
T. Oki[2]
V. Sriboonlue[1]
T. Prempree[1]

[1]Khon Kaen University, Thailand
[2]Institute of Industrial Science, University of Tokyo, Japan

Because rainfall is one of the most significant parameters in a hydrological model, several stochastic models have been developed to analyze and predict the rainfall process. In recent years, wavelet techniques have been widely applied to various water resources research categories because of their time-frequency representation. This study was undertaken to find an alternative method for rainfall prediction by combining the wavelet technique with a conventional autoregressive (AR) model. The 52-year rainfall records of four stations distributed over the northeastern part of Thailand were analyzed by the coupled Wavelet - AR Model (WARM). Two rainfall variables, the number of rainy days and the amount of rainfall were analyzed to obtain the model parameters. Comparing the correlation of the WARM process with the conventional AR process, it is clear that the WARM process provides a better annual rainfall prediction than the simple AR model.

INTRODUCTION

Rainfall is one of the significant parameters in many hydrological models. Rainfall is a complex atmospheric process, which is space and time dependent and it is not easy to predict. Due to the apparent random characteristics of rainfall series, they are often described by a stochastic process. For water resources planning purposes, a long-term rainfall series is required in hydrological and simulation models. There have been many attempts to find the most appropriate method for rainfall prediction. They include, for example, coupling physical, marine, and meteorological or satellite data with a forecasting model, or even applying several techniques such as the artificial neural network 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. This paper introduces a new approach for long-range rainfall prediction of annual series.

Several time series models have been proposed for modeling annual rainfall series such as the autoregressive model (AR) (Yevjevich,1972), the fractional Guassian noise model (Matalas and Wallis, 1971), autoregressive moving-average models (ARMA) (Carlson et al., 1970) and the disaggregation multivariate model (Valencia and Schaake, 1973). In the past decade, wavelet theory has been introduced to signal processing analysis. In recent years, the wavelet transform has been successfully applied to wave data analysis and other ocean engineering applications (Massel, 2001; Teisseire et al., 2002; Huang, 2004). The time-frequency character of long-term climatic data is investigated using the continuous wavelet transform technique (Lau and Weng, 1995; Torrence and Compo, 1997; Mallat, 1998) and wavelet analysis of wind wave measurements obtained from a coastal observation tower (Huang, 2004). With the advantage of the wavelet technique that provides a mathematical process for decomposing a signal into multi-levels of detail, the multiresolution analysis can be done (Liang and Page, 1997). Wavelet transforms were also applied to time series prediction preprocessed for multistep prediction (Tsui et al., 1997). By coupling the wavelet method with the traditional AR model, the Wavelet-Autoregressive model (WARM) is developed for annual rainfall prediction.

DATA AND METHODS

Data input

The point rainfall data of 1951-2002 records from four gauging stations distributed over Northeastern part of Thailand were analyzed by WARM. The selected rainfall stations belong to the Thai Meteorological Department (TMD). Details are shown in Figure 1 and Table 1.

Method and Model structure

Autoregressive model

Autoregressive (AR) models have been extensively applied to hydrology and water resources analysis. The AR has an intuitive type of time dependence, where the value of a variable at the present time depends on the values at previous times. AR models may have constant parameters, parameters varying with time or a 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:

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

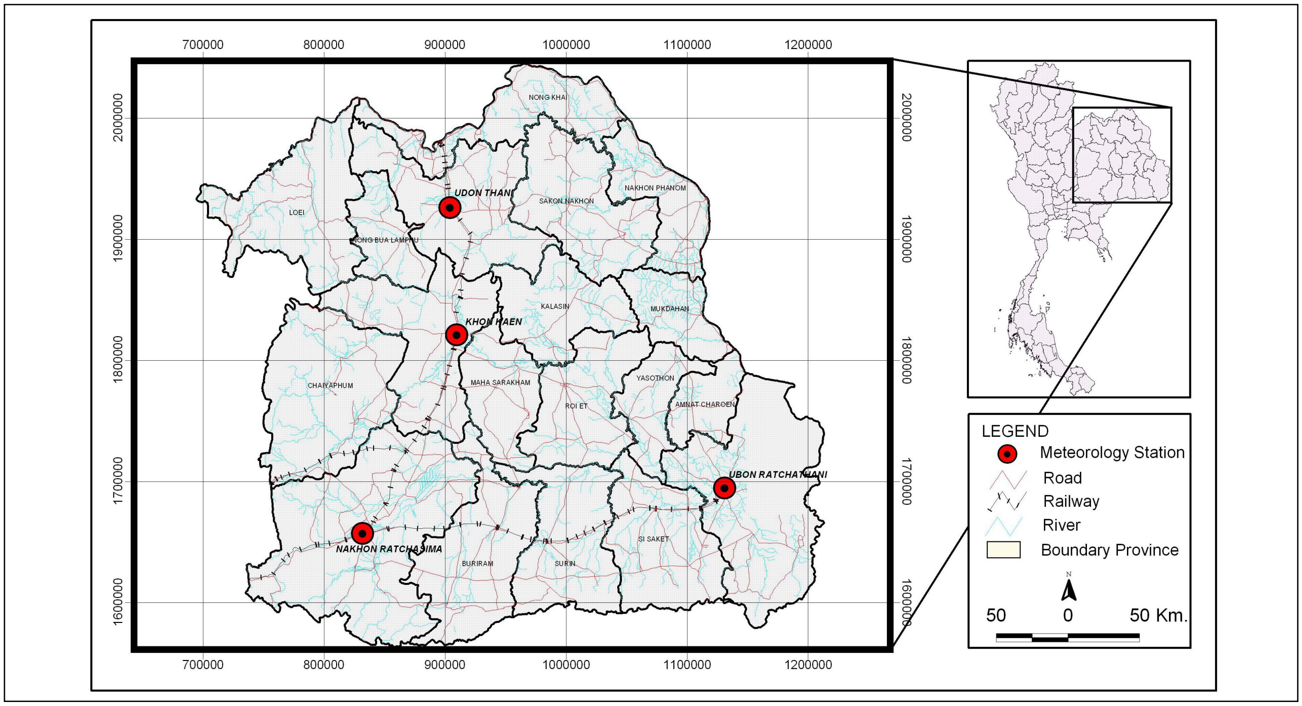

Figure 1. Location of the four selected rainfall stations.

Table 1. Details of Four Selected Rainfall Stations

Station	Period of record	Latitude – longitude	Years of record	Climatic 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

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

where y_t is the time dependent series and ε_t is the time independent series which is uncorrelated with y_t. The series 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 (Fiering and Jackson, 1971; Yevjevich, 1972; Box and Jenkins, 1972) The first order autoregressive (lag-one autoregressive) or first order Markov process can be expressed as:

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

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 a z transform, $z = y - \bar{y}$.

Wavelet transform for signal decomposition

Generally, a signal or function *f(t)* can be expressed in linear decomposition by

$$f(t) = \sum_{l} \alpha_l \psi_l(t) \tag{4}$$

where l is an integer index for the finite or infinite sum, a_l are the real valued expansion coefficients, and $\psi_l(t)$ are the set of real valued functions of t called the expansion set. For the wavelet transform the two parameter system is constructed and Equation (4) becomes

$$f(t) = \sum_{k} \sum_{j} \alpha_{j,k} \psi_{j,k}(t) \tag{5}$$

where j and k are integer indices and $\psi_{j,k}(t)$ are the basis expansion functions of the mother wave $\psi(t)$.

The set of expansion coefficients $\alpha_{j,k}$ are called the discrete wavelet transform (DWT) of *f(t)* and Equation (5) is the inverse transform. The wavelet system is a two-dimensional expansion set (basis) 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

$$\psi_{j,k}(t) = 2^{j/2} \psi\left(2^j t - k\right) \tag{6}$$

where the factor $2^{j/2}$ maintains a constant norm independent of scale j. 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.

In order to generate a set of expansion functions, the signal can be represented by the series

$$f(t) = \sum_{j,k} \alpha_{j,k} 2^{j/2} \psi\left(2^j t - k\right) \tag{7}$$

An efficient way to implement DWT is to use a filter process. Normally, the low frequency content of the signal (approximation, A) is the most important part. It demonstrates the signal identity. The high-frequency component (detail, D) is nuance. The original signal, S, passes through two complementary filters and emerges as two signals of A and D.

Model structure

The concept of the proposed model is based on coupling AR with wavelet transforms. As the wavelet technique can be easily applied for signal analysis, this study used the technique to decompose the details (D) from the approximations (A) of rainfall records (y). In wavelet analysis, the approximations are the high-scale, low frequency components of the signal, and the details are the low-scale, high frequency components. The process of 3-layer decomposition by wavelets is demonstrated in Figure 2 as a wavelet decomposition tree. 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 WARM process can be easily stated as AR with wavelet transform. The schematic diagram, Figure 3, shows the process of WARM. There are two parts of the wavelet transform involved; a

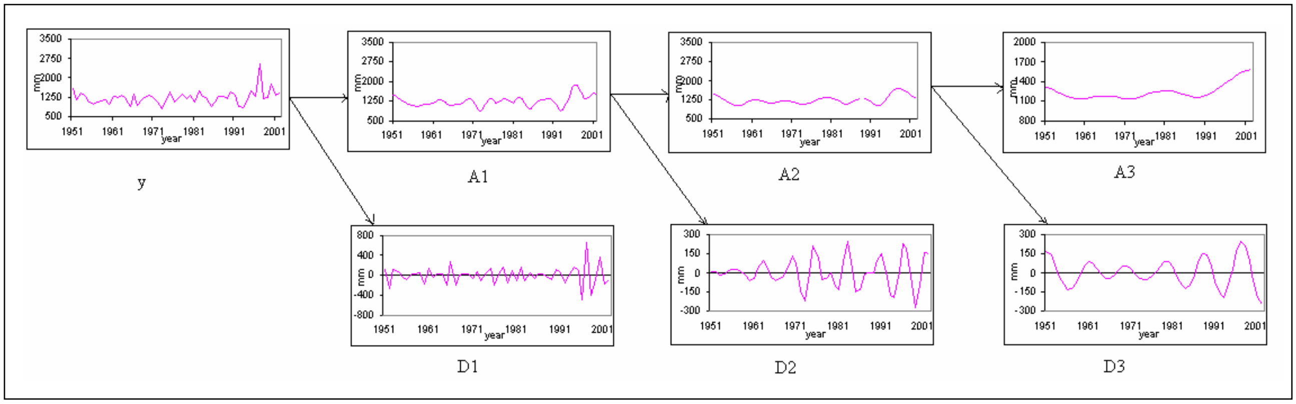

Figure 2. Wavelet decomposition tree of rainfall series at Khon Kaen station. (Ai , Di = Approximation and Detail from i level of filter)

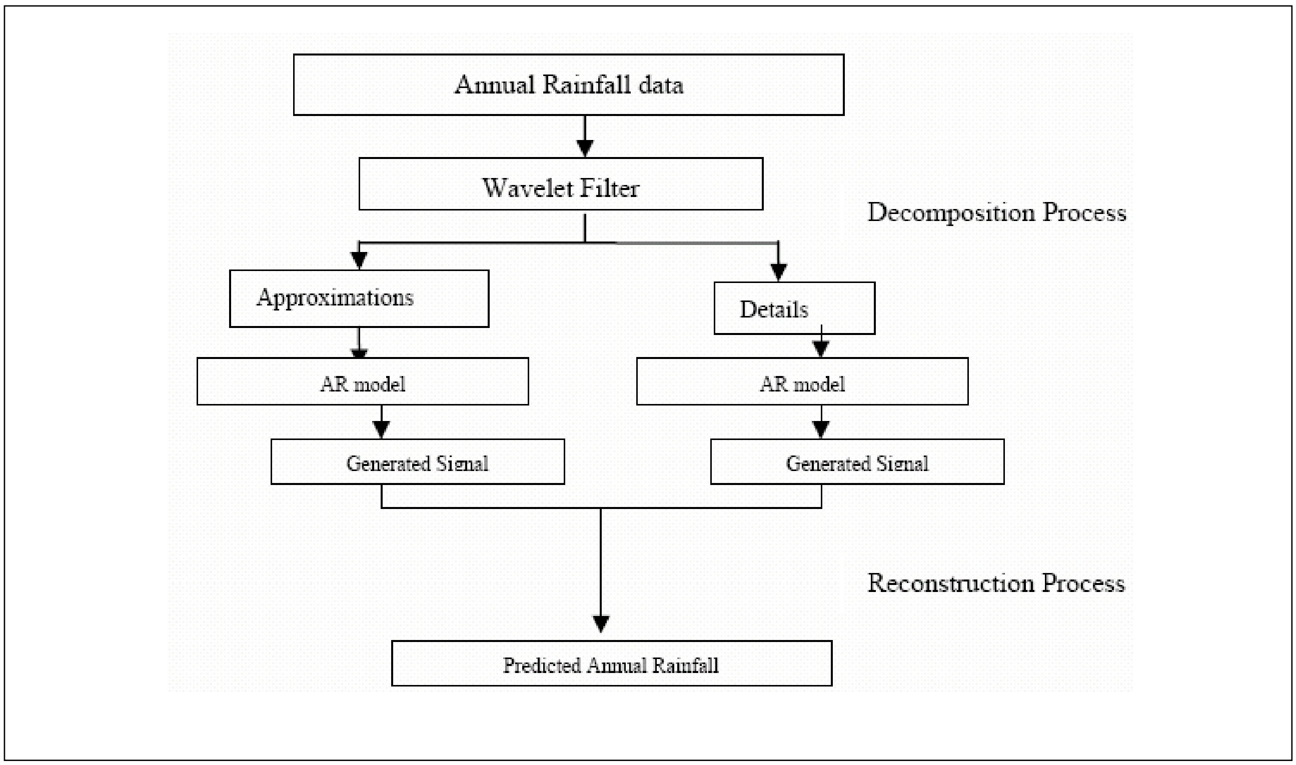

Figure 3. Structure of the WARM process.

decomposition and a reconstruction process. The principle of the model is to split the signal into high frequency and low frequency components. The two components are processed by the AR model for the prediction and then reconstructed into a predicted signal.

RESULTS AND DISCUSSION

The sequences of the number of rainy days (days) and the rainfall (mm) from four selected stations have been analyzed using the WARM model 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. From the simulation result, Figure 4 shows the representatives of each model compared with the historical series. It is apparent that WARM in every level of filtering can represent the characteristics of the selected series for both sequences, whereas the three AR models can only explain trends compared with the previous values. The extreme values cannot be detected by using

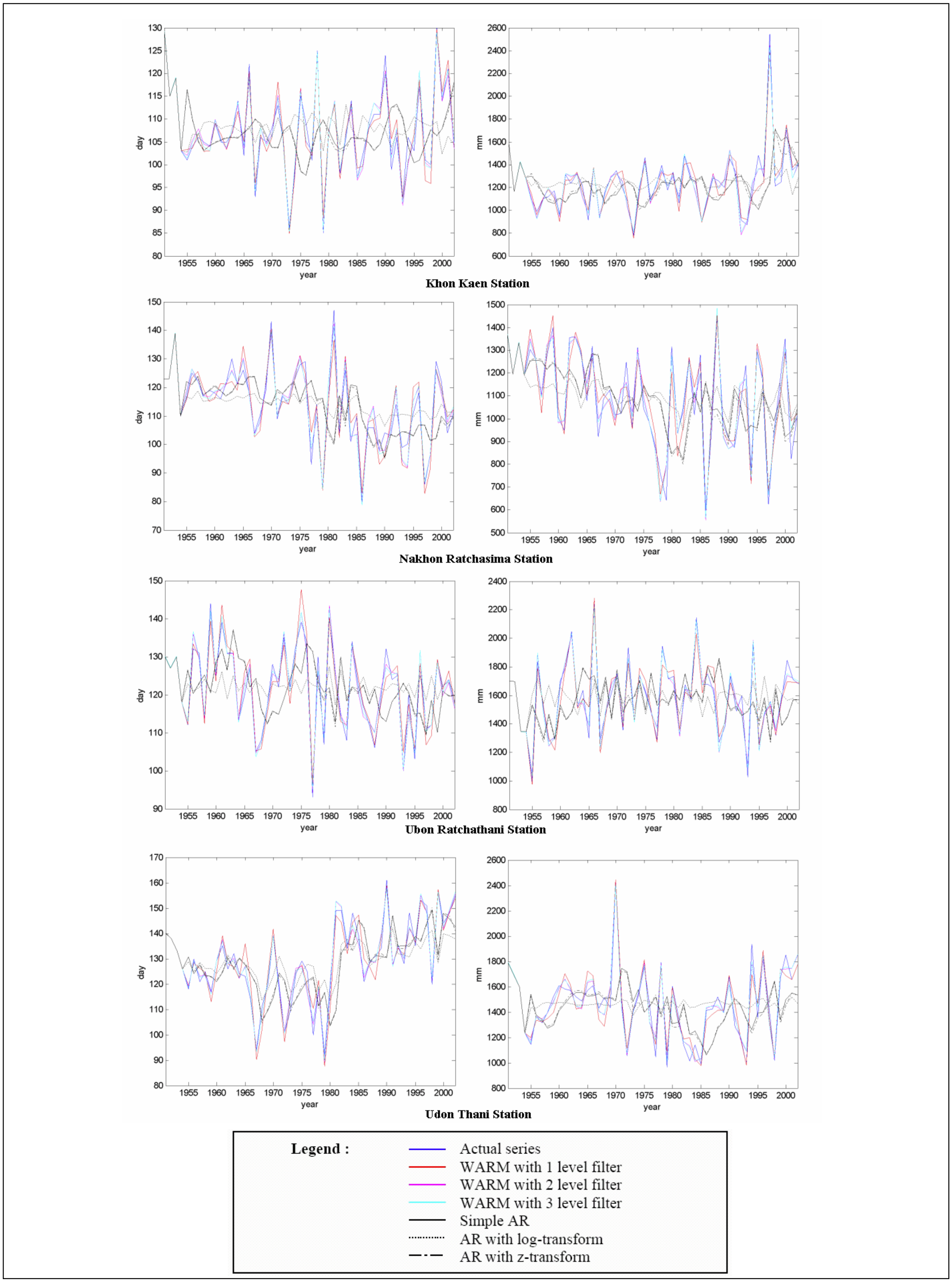

Figure 4 Comparing the studied model with historical data for (1) number of rainy day per year, day; and (2) rainfall, mm

AR or AR with transforms. By comparing the obtained simple R-squared (R^2) and mean square error, it is obvious that WARM with all levels of filter provides better prediction than the traditional AR. The results are shown in Tables 2 and 3. Among the models tested, The WARM with three levels of filter is the most appropriate model for predicting both the number of rainfall days per year and annual rainfall.

Table 2. Obtained R-squared and Mean Square Error from Models for Number of Rainy Days Per Year Series

Model	Udonthani		Khom Kaen		Nakhonratchasima		Ubonratchathani	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
WARM-1 level Filter	0.8983	22.6648	0.9322	5.8708	0.8874	21.0323	0.9294	8.7720
WARM- 2 level filter	0.9646	7.8990	0.9615	3.3350	0.9346	12.2230	0.9626	4.6413
WARM- 3 level filter	0.9693	6.8496	0.9670	2.8596	0.9398	11.2396	0.9684	3.9218
AR model	0.1023	200.1006	0.1454	99.2429	0.0425	194.7934	0.1850	147.2198
AR with z-transformed	0.1980	178.7744	0.2027	69.0807	0.1041	167.3857	0.0415	119.0725
AR with Log-normalized	0.0921	202.3861.	0.1495	99.5946	0.0476	195.7442	0.1961	148.5959

Table 3. Obtained R-squared and Mean Square Error from Models for Annual Rainfall Series

Model	Udonthani		Khom Kaen		Nakhonratchasima		Ubonratchathani	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
WARM-1 level Filter	0.8907	8538.1754	0.9228	5327.7085	0.8954	4554.9483	0.9210	5317.6073
WARM- 2 level filter	0.9346	5110.6815	0.9568	2981.6889	0.9346	2849.9213	0.9633	2466.7780
WARM- 3 level filter	0.9425	4496.4688	0.9603	2741.2951	0.9404	2595.1318	0.9684	2126.6607
AR model	0.1940	93302.7074	0.1632	80277.5997	0.0811	47077.6625	0.1263	75784.0255
AR with z-transformed	0.0416	74890.2783	0.0654	64500.7754	0.0854	39825.8320	0.1106	59843.2656
AR with Log-normalized	0.1925	93183.0017	0.1539	79634.2384	0.1048	48110.5720	0.1321	76170.0873

CONCLUSIONS

The WARM model is an efficient alternative for annual rainfall prediction. It is an improvement over the AR method using the technique of wavelet transforms. The process can be used in place of traditional AR to predict the annual rainfall for a year ahead for the purpose of short-term water resources planning. The model can also be applied to generate long-term annual rainfall series to simulate the water resources system for the purpose of long term planning. Where rainfall prediction is needed in several hydrological models, WARM can be applied to those models in order to make irrigation or water resources planning management more effective.

REFERENCES

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

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

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

Feiring, M.B., and B.B. Jackson. 1971. Synthetic hydrology monograph No.1, American geophysical Union, Washington D.C.

Tsui, F.C., C.C. Li, M. Sun, and R.J. Sclabassi. 1997. A comparative study of two biorthogonal wavelet transforms in time series prediction, Proc. IEEE Int. Conf. Syst. Man. Cybern. Vol. 2, pp. 1791-1796

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

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

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

Matalas, N.C., and J.R. Wallis. 1971. Statistical properties of multivariate fractional noise process, Water Resour. Res., 7, pp 1460-1468

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

Jury, M.R., and J.L. Melice. 2000. Analysis of Durban rainfall and Nile river flow 1871-1999, Theor. Appl. Climatol. 67, pp 161-169

Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane. 1985. Applied modeling of hydrological time series, Book Crafters, Inc. Michigan.

Teisseire, L.M., M.G. Delafoy, D.A. Jordan , R.W. Miksad, and D.C. Weggel. 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, pp 16-24

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

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

Yevjevich, V. 1972. Stochastic processes in hydrology, Water Resources Publications, Colorado.

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