

MATSUYAMA CITY RAINFALL DATA ANALYSIS USING WAVELET TRANSFORM

Celso A. G. SANTOS¹, Carlos de O. GALVÃO², Koichi SUZUKI³
and Ricardo M. TRIGO⁴

¹Member of JSCE, Dr. of Eng., Associate Professor, Dept. of Civil and Environmental Eng., Ehime University
(3-Bunkyo-cho, Matsuyama 790-8577, Japan)

²Dr. of Eng., Associate Professor, Department of Civil Engineering, Federal University of Paraíba
(Av. Aprígio Veloso, 882, Campina Grande, Paraíba, 58109-970, Brazil)

³Member of JSCE, Dr. of Eng., Professor, Dept. of Civil and Environmental Engineering, Ehime University

⁴Ph.D., Senior Research Associate, Department of Physics, University of Lisbon
(Campo Grande. Ed C8, Piso 6, 1749-016, Lisbon, Portugal)

An application of wavelet analysis is done with the total monthly rainfall data of Matsuyama city, in order to analyze the rainfall variability observed in such an area. Besides the rainfall variability analysis, the main frequency components in the time series are studied by the global wavelet spectrum, revealing that the monthly rainfall in Matsuyama city is composed mainly by an annual frequency. Thus, the modulation in the 8–16-month band is examined by an average of all scales between 8 and 16 months, giving a measure of the average monthly variance versus time, where the periods with low or high variance could be identified.

Key Words : *rainfall data, multiscale analysis, wavelet*

1. INTRODUCTION

The wavelet transform is a recent advance in signal processing that has attracted much attention since its theoretical development in 1984 by Grossman and Morlet¹⁾. Its use has increased rapidly as an alternative to the Fourier Transform (FT) in preserving local, non-periodic, multiscaled phenomena. It has advantage over classical spectral analysis, because it allows analyzing different scales of temporal variability and it does not need a stationary series. Thus, it is appropriate to analyze unregular distributed events and time series that contain nonstationary power at many different frequencies. Then, it is becoming a common tool for analyzing localized variations of power within a time series.

Several applied fields are making use of wavelets such as astronomy, acoustics, data compression, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, radar, human vision, pure mathematics, and geophysics such as tropical convection, the El Niño-Southern Oscillation, atmospheric cold fronts, temperature variability, the

dispersion of ocean waves, wave growth and breaking, structures in turbulent flows, and stream flow characterization.^{2),3),4),5)}

The following sections describe the wavelet transform, the rainfall data of Matsuyama city, and then the application of wavelet to such data using the program developed by Torrence and Compo.³⁾

2. WAVELET TRANSFORM

Mathematical transformations are applied to signals to obtain further information from that signal that is not readily available in the raw signal. There are several transformations that can be applied, among which the Fourier transforms are probably by far the most popular. In order to maintain time and frequency localization in a signal analysis, one possibility would be to do a Windowed Fourier Transform (WFT), using a certain window size and sliding it along in time, computing the Fast Fourier Transform (FFT) at each time using only the data within the window. This would solve the frequency localization problem, but would still be dependent on the window size used. The main problem with

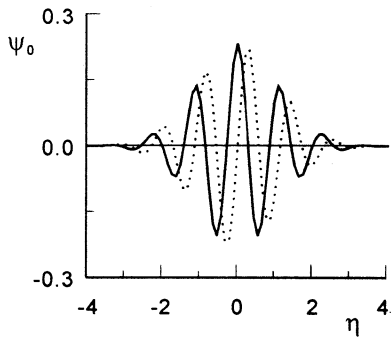


Fig. 1 Morlet wavelet base with frequency $\omega_0 = 6.0$, in which the real part is in solid line and imaginary part is in dashed line.

the WFT is the inconsistent treatment of different frequencies: at low frequencies there are so few oscillations within the window that the frequency localization is lost, while at high frequencies there are so many oscillations that the time localization is lost. Finally, the WFT relies on the assumption that the signal can be decomposed into sinusoidal components.

Thus, to measure the stationarity of a time series is necessary to calculate the running variance using a fixed-width window. Despite the disadvantage of using a fixed-width window, the analysis could be repeated with a variety of window widths. By smoothly varying the window width, a picture of the changes in variance versus both time and window width could be built. The obvious problem with this technique is the simple “boxcar” shape of the window function, which introduces edge effects such as ringing. Using such a black-box-car, there will be no information on what is going on within the box, but only recover the average energy.

Wavelet analysis attempts to solve these problems by decomposing or transforming a one-dimensional time series into a diffuse two-dimensional time-frequency image simultaneously. Then, it is possible to get information on both the amplitude of any “periodic” signals within the series, and how this amplitude varies with time.

An example of a wave “packet”, of finite duration and with a specific frequency, is shown in Fig. 1. Such a shape could be used as a window function for the analysis of variance. This “wavelet” has the advantage of incorporating a wave of a certain period, as well as being finite in extent. In fact, the wavelet shown in Fig. 1 (called the Morlet wavelet) is nothing more than a Sine wave multiplied by a Gaussian envelope.

Assuming that the total width of this wavelet is about 10 years, it is possible to find the correlation between this curve and the first 10 years of the time series later shown in Fig. 3a. This single number gives a measure of the projection of this wave

packet on the data during the 1890-1900 period, i.e. how much [amplitude] does the 10-year period resemble a Sine wave of this width [frequency]. By sliding this wavelet along the time series, a new time series of the projection amplitude versus time can be constructed.

Finally, the “scale” of the wavelet can be varied by changing its width. This is the real advantage of wavelet analysis over a moving Fourier spectrum. For a window of a certain width, the sliding FFT is fitting different numbers of waves; i.e., there can be many high-frequency waves within a window, while the same window can only contain a few (or less than one) low-frequency waves. The wavelet analysis always uses a wavelet of the exact same shape, only the size scales up or down with the size of the window.

In addition to the amplitude of any periodic signals, it is worth to get information on the phase. In practice, the Morlet wavelet shown in Fig. 1 is defined as the product of a complex exponential wave and a Gaussian envelope:

$$\Psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2} \quad (1)$$

where $\Psi_0(\eta)$ is the wavelet value at nondimensional time η , and ω_0 is the nondimensional frequency, equal to 6 in this study in order to satisfy an admissibility condition; i.e., the function must have zero mean and be localized in both time and frequency space to be “admissible” as a wavelet. This is the basic wavelet function, but it will be now needed some way to change the overall size as well as slide the entire wavelet along in time. Thus, the “scaled wavelets” are defined as:

$$\Psi\left[\frac{(n' - n)\delta t}{s}\right] = \left(\frac{\delta t}{s}\right)^{1/2} \Psi_0\left[\frac{(n' - n)\delta t}{s}\right] \quad (2)$$

where s is the “dilation” parameter used to change the scale, and n is the translation parameter used to slide in time. The factor of $s^{-1/2}$ is a normalization to keep the total energy of the scaled wavelet constant.

We are given a time series X , with values of x_n , at time index n . Each value is separated in time by a constant time interval δt . The wavelet transform $W_n(s)$ is just the inner product (or convolution) of the wavelet function with the original time series:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \Psi^* \left[\frac{(n' - n)\delta t}{s} \right] \quad (3)$$

where the asterisk (*) denotes complex conjugate.

The above integral can be evaluated for various values of the scale s (usually taken to be multiples of the lowest possible frequency), as well as all values of n between the start and end dates. A two-dimensional picture of the variability can then be constructed by plotting the wavelet amplitude and

phase. Then, a time series can be decomposed into time-frequency phase space using a typical (mother) wavelet. The actual computation of the wavelet transform can be done by the following algorithm³⁾: (a) choose a mother wavelet; (b) find the FT of the mother wavelet; (c) find the FT of the time series; (d) choose a minimum scale s_0 , and all other scales; (e) for each scale, do:

- Using Eq. (4), or whatever is appropriate for the mother wavelet in use, compute the daughter wavelet at that scale:

$$\Psi(s\omega_k) = \left(\frac{2\pi s}{\delta t} \right)^{1/2} \hat{\Psi}_0(s\omega_k) \quad (4)$$

where the $\hat{\cdot}$ indicates the FT.

- Normalize the daughter wavelet by dividing by the square-root of the total wavelet variance (the total of Ψ^2 should then be one, thus preserving the variance of the time series);
- Multiply by the FT of your time series;
- Using Eq. (5), inverse transform back to real space;

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\Psi}^*(s\omega_k) e^{i\omega_k n \delta t} \quad (5)$$

where ω_k is the angular frequency, equal to $2\pi k/N\delta t$ for $k \leq N/2$ or equal to $-2\pi k/N\delta t$ for $k > N/2$. It is possible to compute the wavelet transform in the time domain using Eq. (3). However, it is much simpler to use the fact that the wavelet transform is the convolution between the two functions x and Ψ , and to carry out the wavelet transform in Fourier space using the FFT; and (f) make a contour plot.

3. RAINFALL DATA

Matsuyama city lies in the middle of Ehime Prefecture, Japan, at $33^\circ 50'$ N latitude and $132^\circ 46'$ E longitude. Its area is 289.3 km^2 stretching 30.2 km from east to west, and 28.7 km from north to south. It is about 650 km from Tokyo and about 420 km from Osaka. To the east of Matsuyama lies Mt. Ishizuchi ($1,982 \text{ m}$), the highest mountain in western Japan. Matsuyama faces the calm Seto Inland Sea, and the Shigenobu and the Ishite rivers flow east to west through the city.

Matsuyama is blessed with natural surroundings, has very few natural disaster, and little snow in winter. Its population has increased to $452,000$, which is more than 13 times as large as when the city was founded in 1889. This area is part of the Seto Inland Sea climate region, which is mild with little rainfall throughout the four seasons and the

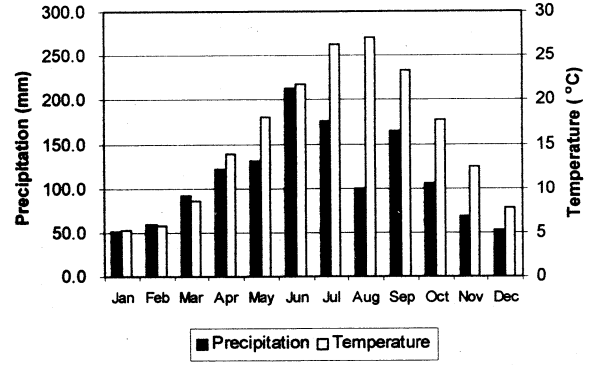


Fig. 2 Mean precipitation and temperature in Matsuyama city.

annual mean temperature is 15.3°C , with an annual mean rainfall equal to $1,333 \text{ mm}$, based on the record of the last 110 years. The rainy season is from middle of June to middle of July and the typhoons mainly occur between August and October. The monthly mean precipitation and temperature are given in Fig. 2 and the total monthly rainfall series since 1890 to May 2000 is shown in Fig. 3a. These weather data have been recorded by Matsuyama Weather Station.

4. DATA ANALYSIS

Wavelet analysis was chosen, besides the other reasons and advantages described here, because applications such as standard Fourier Transform analysis to a time series should be only attempted when the time series fulfils two important characteristics, namely: (1) stationarity; i.e., that no changes in the mean, variance, etc, occur throughout the time series; and (2) that the time series can be described as the summation of different periodic components (described by simple harmonic functions) for the whole period. However, most time series from meteorology and hydrology do not fulfil both requirements. In fact, earth sciences time series are usually nonstationary and present trends of the mean value, changes in the variability for certain periods. Furthermore, many hydrological time series, such as precipitation, present unregular distributed events with nonstationary power over many different frequencies. Thus, their intrinsic temporal structure is not well represented by the superposition of a few frequency components as derived in a usual Fourier analysis.

(1) Wavelet power spectrum

Since the present data are monthly distributed, the parameters for the wavelet analysis are set as $\delta t = 1$ month and $s_0 = 2$ months because $s = 2\delta t$, $\delta j = 0.25$ to do 4 sub-octaves per octave, and $j_1 = 7/\delta j$ in order

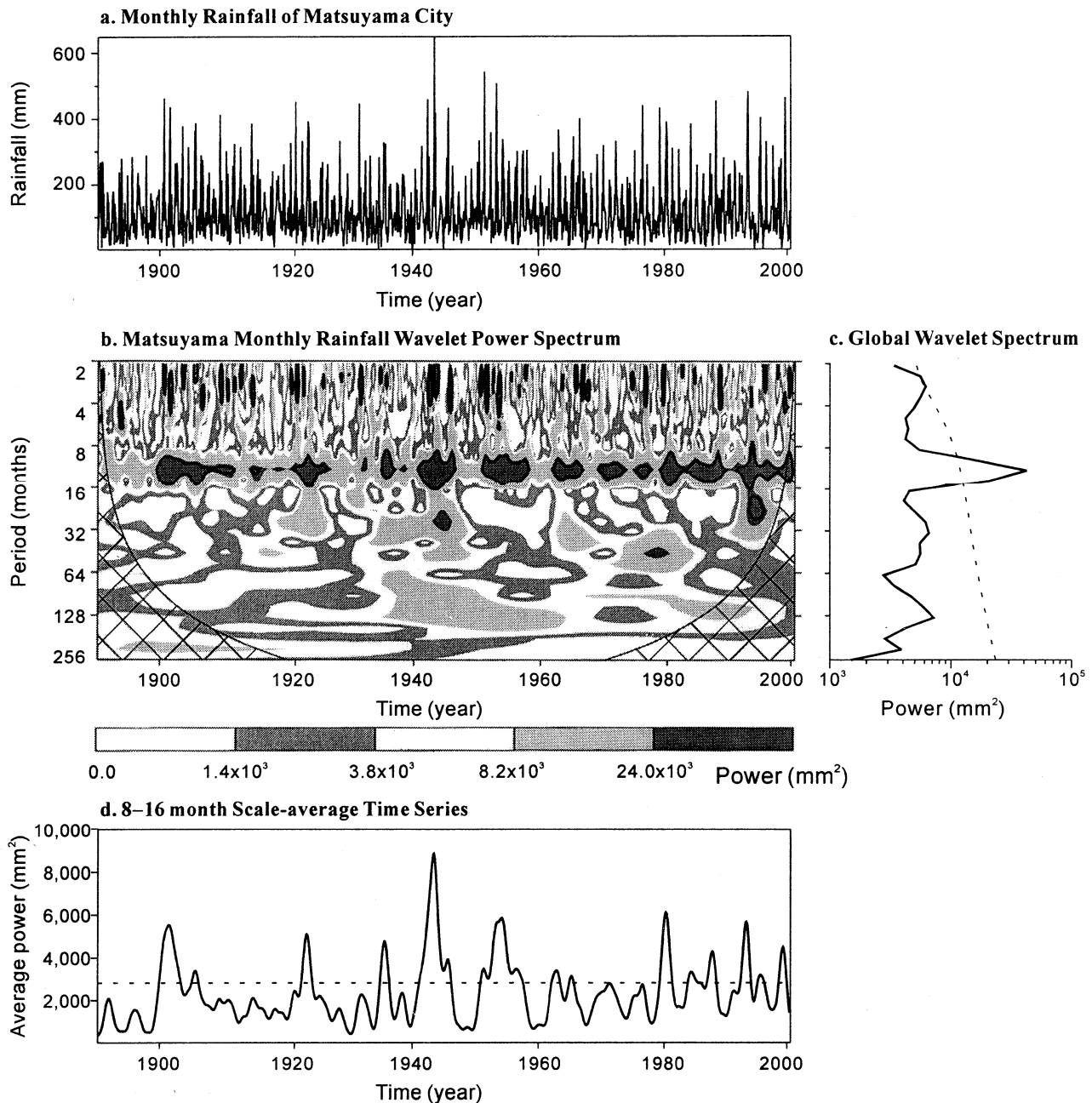


Fig. 3 (a) Total monthly rainfall series of Matsuyama city. (b) The wavelet power spectrum. The contour levels are chosen so that 75%, 50%, 25%, and 5% of the wavelet power is above each level, respectively. Cross-hatched region is the cone of influence, where zero padding has reduced the variance. Black contour is the 5% significance level, using a red-noise ($\alpha = 0.2787$) background spectrum. (c) The global wavelet power spectrum (black line). The dashed line is the 5% significance level for the global wavelet spectrum. (d) Scale-average wavelet power over the 8–16-month band for the total monthly rainfall in Matsuyama. The dashed line is the 95% confidence level assuming red noise $\alpha = 0.2787$.

to do 7 powers-of-two with δj sub-octaves each.

Figure 3b shows the power (absolute value squared) of the wavelet transform for the monthly rainfall in Matsuyama city presented in **Fig. 3a**, which is a record of the last 110 years. As stated before, the $(\text{absolute value})^2$ gives information on the relative power at a certain scale and a certain time. This figure shows the actual oscillations of the individual wavelets, rather than just their magnitude. Observing **Fig. 3b**, it is clear that there is more

concentration of power between the 8–16-month band, which shows that this time series has a strong annual signal. Classical statistical analysis applied by previous authors for this area has mentioned the existence of important low frequency peaks. Here, we show that such results are misleading as no significant peaks were attained for low frequency periods (e.g., 32 and 128 months). However, wavelet power spectrum for precipitation episodes with characteristic scale of 2–4-month presents an

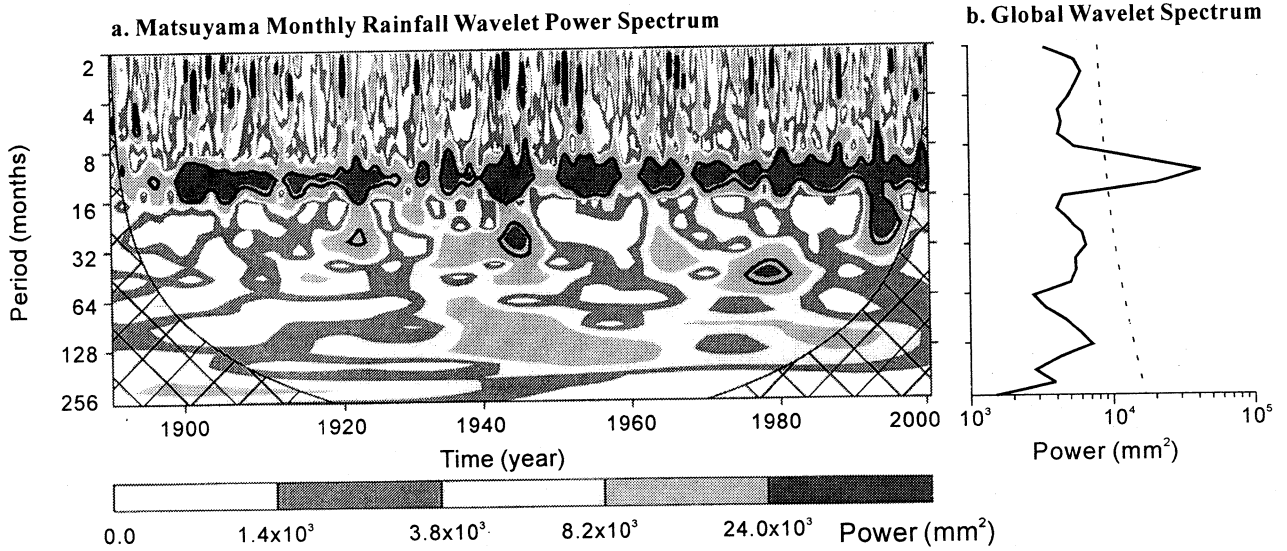


Fig. 4 (a) The Matsuyama monthly rainfall wavelet power spectrum. The contour levels are chosen so that 75%, 50%, 25%, and 5% of the wavelet power is above each level, respectively. Cross-hatched region is the cone of influence, where zero padding has reduced the variance. Black contour is the 5% significance level, using a white-noise background spectrum. (b) The global wavelet power spectrum (black line). The dashed line is the 5% significance level for the global wavelet spectrum, using white-noise background spectrum.

important peak, almost significant at the 5% level. The variance of power in 8–16-month band (also confirmed later by **Fig. 3d**) also shows the dry and wet years; i.e., when the power decreases substantially in this band, it means a dry year and when the power is maximum means a wet year. For example, a dry period can be identified during late XIX century, until the beginning of 1900's followed by a wet period until the beginning of 1910's. An extreme reduction in power can be found between the years 1935 and 1945 in the year 1939, which corresponds to a dry year. Another wet period that contains certain reductions can be identified since 1979 whose fast reductions of the power are also due to the presence of very singular dry years, e.g., years 1978 and 1994.

The cross-hatched region in this figure is the cone of influence, where zero padding has reduced the variance. Because we are dealing with finite-length time series, errors will occur at the beginning and end of the wavelet power spectrum. One solution is to pad the end of the time series with zeroes before applying the wavelet transform and then remove them afterward. Here the time series is padded with sufficient zeroes to bring the total length N up to the next-higher power of two, thus limiting the edge effects and speeding up the Fourier Transform. Padding with zeroes introduces discontinuities at the endpoints and decreases the amplitude near the edges as going to larger scales, since more zeroes enter the analysis. The cone of influence is the region of the wavelet spectrum in which edge effects become important and is defined as the e-

folding time for the autocorrelation of wavelet power at each scale. The peaks within these regions have presumably been reduced in magnitude due to the zero padding. Thus, it is unclear whether the decrease in any band power in this cross-hatched region is a true decrease in variance or an artifact of the padding. For much narrower mother wavelets such as Mexican hat wavelet their cone of influence would be much smaller and thus is less affected by edge effects. Note also that for cyclic series, there is no need to pad with zeroes, and there is no cone of influence.

The black contour in the same figure is the 5% significance level, using a red-noise background spectrum. Many geophysical time series can be modeled as either white-noise (**Fig. 4**) or red-noise (**Fig. 3**). A simple model for red-noise is the univariate lag-1 autoregressive process. The lag-1 is the correlation between the time series and itself, but shifted (or lagged) by one time unit. In this present case, this would be a shift of one month. The lag-1 measures the persistence of an anomaly from one month to the next. The true lag-1 α can be computed by an approximation using $\alpha = (\alpha_1 + \alpha_2^{1/2})/2$, where α_1 is the lag-1 autocorrelation and α_2 is the lag-2 autocorrelation, which is the same as lag-1 but just shifted by two points instead of one. Since the present time series shows an $\alpha_1 = 0.1983$ and $\alpha_2 = 0.1289$, the true lag-1 α is assumed to be 0.2787. Results equivalent to those presented in **Figs 3b** and **3c** are shown in **Fig. 4** but with the 5% significance level using a white-noise background spectrum.

Note that since the present time series has a small value of red-noise autocorrelation, the black contours in **Fig. 4a** are similar to the contours in **Fig. 3b**.

The null hypothesis is defined for the wavelet power spectrum as assuming that the time series has a mean power spectrum; if a peak in the wavelet power spectrum is significantly above this background spectrum, then it can be assumed to be a true feature with a certain percent confidence. For definitions, "significant at the 5% level" is equivalent to "the 95% confidence level," and implies a test against a certain background level, while the "95% confidence interval" refers to the range of confidence about a given value. The 95% confidence implies that 5% of the wavelet power should be above this level.

(2) Global wavelet power spectrum

The annual frequency (periodicity at 12 months) of this time series is confirmed by an integration of power over time (**Figs 3c** and **4b**), which shows only one significant peak above the 95% confidence level for the global wavelet spectrum, assuming $\alpha = 0.2787$ (**Fig. 3c**) or assuming white-noise (**Fig. 4b**), represented by the dashed lines. However, **Fig. 3c** also presents an almost significant peak (at the 5% level) centered in the 2–4-month band. In fact, most extreme monthly precipitation values for Matsuyama city (values above 300 mm in **Fig. 3a**) correspond to pulses of highly significant power within the 2–4-month band (**Fig. 3b**). This global wavelet spectrum provides an unbiased and consistent estimation of the true power spectrum of the time series, and thus it is a simple and robust way to characterize the time series variability. Global wavelet spectra should be used to describe rainfall variability in non-stationary hyetographs. For regions that do not display long-term changes in hyetograph structures, global wavelet spectra are useful for summarizing a region's temporal variability and comparing it with rainfall in other regions. The global wavelet spectral shape is controlled primarily by the distribution of feature scales, and appears diagnostic of the hydroclimatic regime despite a large range in watershed sizes because a clear qualitative difference could be found in the global wavelet spectra of hyetographs from different climatic regions.

(3) Scale-average time series

The scale-average wavelet power (**Fig. 3d**) is a time series of the average variance in a certain band, in this case 8–16-month band, used to examine modulation of one time series by another, or modulation of one frequency by another within the

same time series. This figure is made by the average of **Fig. 3b** over all scales between 8 and 16 months, which gives a measure of the average year variance versus time. The variance plot shows distinct periods when monthly rainfall variance was low, e.g., from 1890 to 1900 and from 1910 to 1920, and an important peak of power spectrum can be identified for 1993, clearly indicating wetter than normal period.

5. CONCLUSION

In order to study the variability of the monthly rainfall time series in Matsuyama city, wavelet analysis is applied. The wavelet power spectrum shows a big power concentration between the 8–16-month band, revealing an annual periodicity of such events, which is confirmed by the peak of the integration of transform magnitude vectors over time that show again a strong annual signal. The periods with low variance in such a band can be identified by the average of the all scales between 8 and 16 months, which gives a measure of the average monthly variance versus time. Further study could be done with wavelet analysis of rainfall in Matsuyama area together with stream flow analysis in order to benefit erosion models from a quantitative breakdown of the temporal components of stream flow, particularly for Ishite river watershed, in which sediment supply is limited.

ACKNOWLEDGMENT: The writers wish to thank Dr. Christopher Torrence of Advanced Study Program at National Center for Atmospheric Research, Colorado, for providing the wavelet analysis computer program and material, as well as for his comments.

REFERENCES

- 1) Grossman, A. and Morlet, J.: Decomposition of Hardy functions into square integrable wavelets of constant shape, *SIAM J. Math. Anal.*, **15**, 723-736, 1984.
- 2) Graps, A.: An introduction to wavelets, *IEEE Computational Science and Engineering*, **2**, No. 2, 50-61, 1995.
- 3) Torrence, C. and Compo, G.P.: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.*, **79**, No. 1, 61-78, 1998.
- 4) Farge, M.: Wavelet transforms and their applications to turbulence. *Ann. Rev. Fluid Mech.*, **24**, 395-457, 1992.
- 5) Smith, L.C., Turcotte, D.L. and Isacks, B.L.: Stream flow characterization and feature detection using a discrete wavelet transform, *Hydrological Processes*, **12**, 233-249, 1998.

(Received October 2, 2000)