Predictions in Ungauged Basins: Promise and Progress (Proceedings of symposium S7 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005). IAHS Publ. 303, 2006.

Application of the wavelet transform for analysis of precipitation and runoff time series

CELSO AUGUSTO G. SANTOS[1] &
SANDRA MARIA ARAÚJO IDEIÃO[2]

1 *Department of Civil Engineering, Federal University of Paraíba, 58051-900, João Pessoa-PB, Brazil*
celso@ct.ufpb.br

2 *PIBIC/CNPQ Scholar, Department of Civil Engineering, Federal University of Paraíba, Brazil*

Abstract In order to help the hydrological regionalization, the analysis of precipitation and runoff data was done through the wavelet transform. This analysis seeks to help water administration personnel to take decisions on Predictions in Ungauged Basins, providing more details concerning the information about the precipitation and runoff patterns within a region. Data from rainfall and runoff stations in Piranhas-Açu River basin, located in semiarid northeastern Brazil, were used and the wavelet transforms were applied to these time series in order to determine zones within the region. In spite of the rainfall wavelet power spectra showed to be very similar, the global wavelet spectra of the main frequency components of the stream flow time series revealed different patterns. A total of 12 rainfall and runoff time series were assessed and their global wavelet spectra, together with the band scale-average time series, can be considered useful for determining the hydrological zones within a region.

Key words rainfall data; runoff data; multiscale analysis; wavelet

INTRODUCTION

The wavelet transform is a recent advance in signal processing that has attracted much attention since its theoretical development in 1984 by Grossman & Morlet (1984). Its use has increased rapidly as an alternative to the Fourier transform in preserving local, non-periodic, multiscaled phenomena, and it has advantages over classical spectral analysis, because it allows analysis of different scales of temporal variability and it does not need a stationary series (Smith *et al.*, 1998). Thus, it is appropriate to analyse irregular distributed events and time series that contain nonstationary power at many different frequencies. Several applied fields are making use of wavelets, such as astronomy, acoustics, data compression, and nuclear engineering (Farge, 1992; Graps, 1995; Torrence & Compo, 1998). Although, the application of wavelet transform is not frequently used in hydrology, their use is also increasing, e.g. to help basin characterization (Smith *et al.*, 1998; Gaucherel, 2002) and in the study of hydrological regime variability (Labat *et al.*, 2004).

Hydrological regionalization, i.e. information transfer to sites without flow records using the available flow records within the same region, is usually done for mean annual flow or mean annual flood, intended to be used for design or planning purposes. Recently, however, there is an increasing demand for information to be used for operation and management of water systems. Instead of long-range averages, the need now is for information, produced in real time, at the space and time variability of the hydrological regime.

Piranhas-Açu River, located in semiarid northeastern Brazil, flows through the States of Paraíba and Rio Grande do Norte. The management of its waters is one problem that demands such information, due to the usually high potential for conflicts and to the frequent difference in monitoring information availability among the several sub-basins involved. Water management between these two Federal States is a difficult task, due to the low available stream flow. Hydrological regionalization is a necessary tool in the basin, in order to help the decision making process. However, regionalization is complicated because the rivers, although located within the same basin, can present different stream flow characteristics, not easily detectable (with standard statistical tools) due to the similarity of the hydrological regime. Thus, the application of wavelet analysis is done in order to help determining the hydrological zones within the region.

Meteorological forecasts with reasonable skill offer useful information of rainfall estimation with antecedence ranging from some days to one year. The operational hydrology challenge is the estimation of seasonal, monthly or daily discharge in an ungauged basin, taking advantage of the rainfall forecasts, and also using available regional hydrological information. One major problem for performing these estimates is that the flow record includes signals of different frequencies, caused by diverse geophysical phenomena. The following sections describe the wavelet transformation, the selected data, and then the application of wavelet to such data using the program developed by Torrence & Compo (1998) in order to use the global wavelet power spectra in the hydrological regionalization process.

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.

Wavelet analysis maintains time and frequency localization in a signal analysis by decomposing or transforming a one-dimensional (1-D) time series into a diffuse 2-D 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 basic wave or mother wavelet, as it is known in the literature, is the Morlet wavelet. This wavelet has the advantage of incorporating a wave of a certain period, as well as being finite in extent. Assuming, for example, 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. This single number gives a measure of the projection of this wave packet on the data during the period, i.e. how much amplitude does this 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 *vs* time can be constructed. Finally, the scale of the wavelet can be varied by changing its width. In continuous time, but on a finite interval, the Morlet wavelet is defined as the product of a complex exponential wave and a Gaussian envelope:

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

where $\Psi_0(\eta)$ is the wavelet value at non-dimensional time η, and ω_0 is the non-dimensional frequency, equal to 6 in this study in order to satisfy the 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 now some way will be needed to change the overall size as well as slide the entire wavelet along in time. Thus, the scaled wavelets are defined as:

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

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:

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

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 2-D 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 mother wavelet.

A significance level can be drawn in the same figure, using a red-noise background spectrum. Many geophysical time series can be modelled as either white-noise or red-noise. 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 or one day. The lag-1 measures the persistence of an anomaly from one month (or day) 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 time units instead of one.

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. More details can be found in Santos *et al.* (2001).

SELECTED DATA

Northeastern Brazil has an area of 1 552 619.2 km^2. In climatic terms, northeastern Brazil can be considered as a complex area, not due to the variation in the temperatures, but due to the variation of the rainfall amounts and distribution. The mean temperatures vary between 23 and 27°C, with minimum temperatures during winter around 5 and 10°C and maximum ones in summer between 30 and 40°C. In order to apply wavelet analysis to the identification of similarities and differences among the hydrological regime of the basins, 12 sub-basins with available mean daily runoff and total monthly rainfall records were selected from the so-called Sub-basin 37, part of the national hydrological network (Fig. 1). The selected sub-basins are Angicos, Upanema, Augusto Severo, Pedra de Abelhas, São Fernando, Sítio Volta, Pau dos Ferros, Serra Negra do Norte in Rio Grande do Norte State, and Antenor Navarro, Aparecida, Piancó and Emas in Paraíba State. Although the rainfall time series ranges from 15 to 76 years, the duration of the runoff time series are variable, unfortunately, ranging from 6 to 40 years. Variables describing the above sub-basins are available, e.g. catchment areas, channel slopes and length of channels.

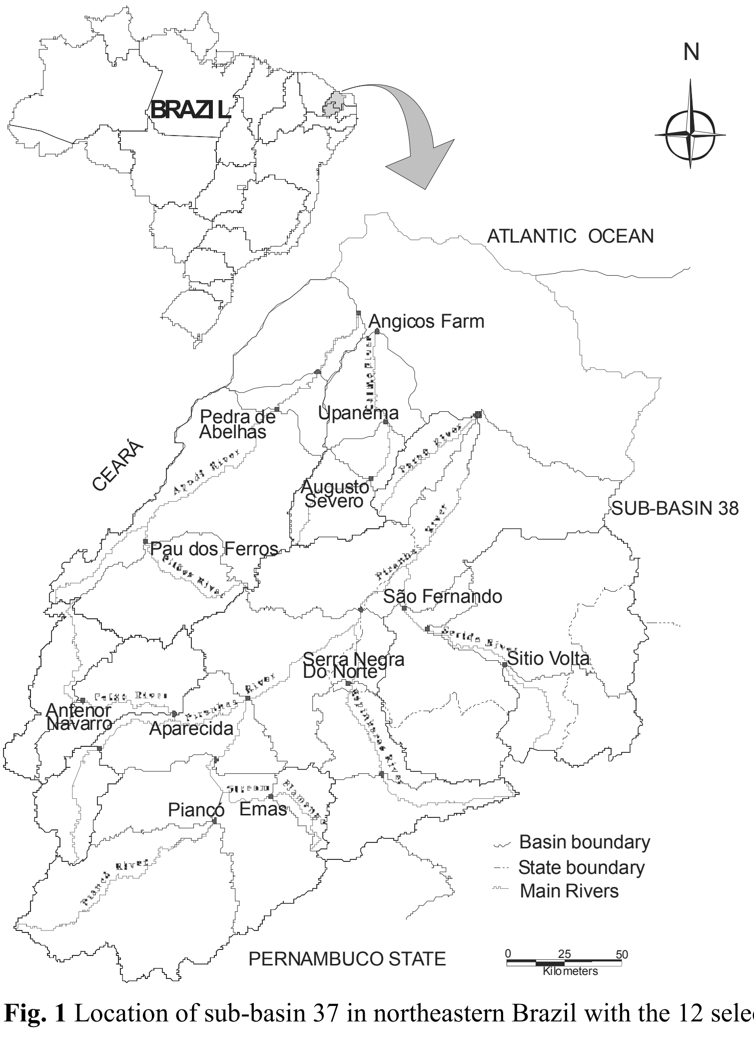

Fig. 1 Location of sub-basin 37 in northeastern Brazil with the 12 selected sub-basins.

DATA ANALYSIS

Total monthly rainfall and mean daily runoff discharge data are used for the analysis. The parameters for the rainfall wavelet analysis are set as the time interval δt = 1 month, the start scale s_0 = 2 months, the scale width δj = 0.25, which will do 4 sub-octaves per octave, and there will be 9 powers-of-two, and for the runoff wavelet analysis they are set as δt = 1 day, s_0 = 2 days, δj = 0.25, and there will be 11 powers-of-two (2, 4, 8, 16, 32, 64, ..., 2048).

Wavelet power spectrum

Since all data from the selected rainfall raingauges showed similar wavelet power spectra, only the case of Angicos sub-basin is depicted herein. Figure 2(a) shows the raw data of the total monthly rainfall at Angicos raingauge from 1911 to 1987 and Fig. 2(b) shows the power (absolute value squared) of the wavelet transform for the total monthly rainfall. The (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 this figure, the

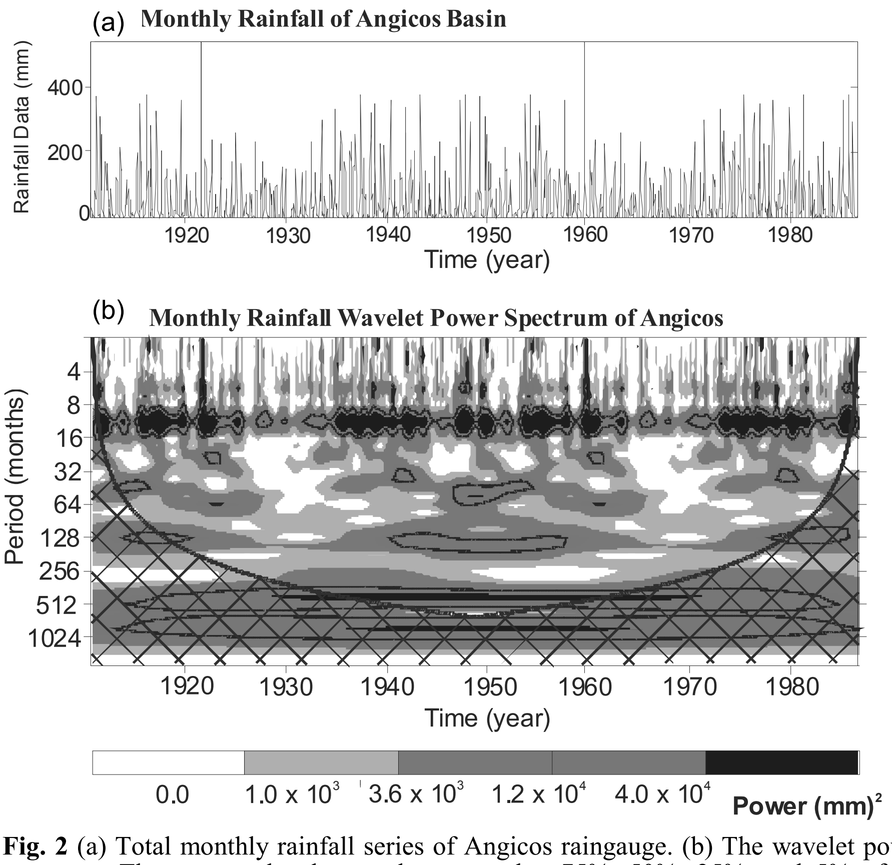

Fig. 2 (a) Total monthly rainfall series of Angicos raingauge. (b) The wavelet power spectrum. The contour levels are chosen so that 75%, 50%, 25%, and 5% of the wavelet power is above each level, respectively. Cross-hatched region is the cone of influence, where zero padding has reduced the variance. Black contour is the 5% significance level, using a white-noise (α = 0.0) background spectrum.

concentration of power can be easily identified in the frequency or time domain. For example, an annual frequency can be observed from 1911 to 1987 with power reduction from 1925 to 1935 and from 1965 to 1975, and more two frequency components are presented, one is a semidecadal frequency from 1945 to 1955 and another one is a 32-year frequency from 1940 to 1960. 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 (Santos *et al.*, 2001).

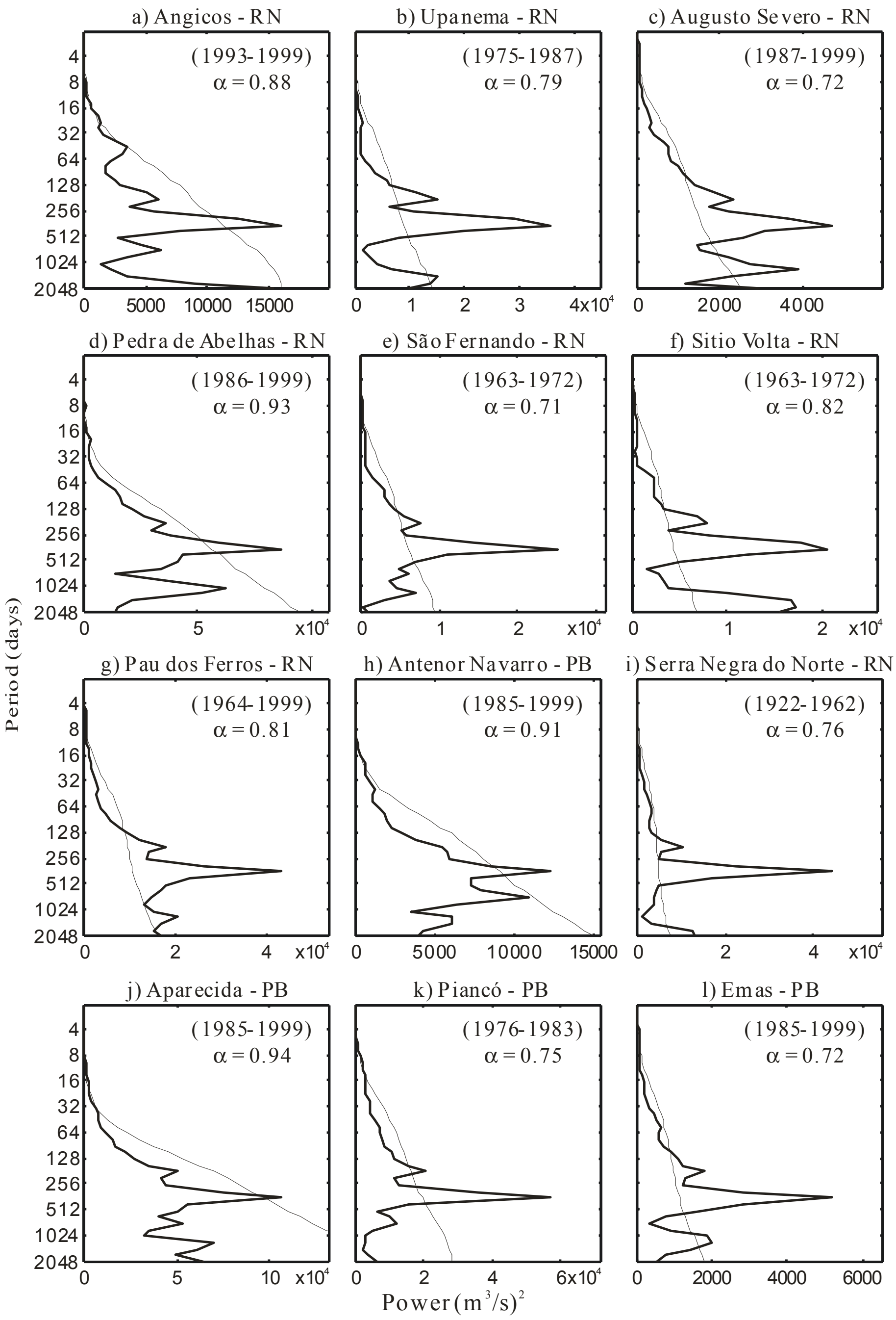

Fig. 3 Global wavelet power spectra for the 12 selected sub-basins.

The lag-1 for all rainfall wavelet analyses are very close to 0.4, then the time series are modelled as white-noise. However, for the runoff wavelet analysis, the lag-1 are much larger than 0.4 and, thus, these time series are modelled as red-noise, as indicated later in each figure.

Global wavelet power spectrum

Global wavelet power spectra are obtained by the time-average of power over time and together with the confidence level, they can confirm the main component frequencies of the time series. The studies rainfall time series showed annual frequencies (periodicity at 12 months) and the main component frequencies of the runoff data are confirmed by this time-average of power over time (Fig. 3), which shows the significant peaks above the 95% confidence level for the global wavelet spectrum, assuming red-noise, represented by the dashed lines. These global wavelet spectra provide 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 the runoff variability in non-stationary hydrographs. For regions that do not display long-term changes in hydrograph structures, global wavelet spectra could be useful for summarizing a region's temporal variability and comparing them with runoff in other regions.

Scale-average time series

The scale-average wavelet power is a time series of the average variance in a certain band. In the case of Figs 4 and 5, they are the semi-annual, annual and biennial bands for the Piancó sub-basin in Paraíba State and Serra Negra do Norte sub-basin in Rio Grande do Norte Sate, respectively. The scale-average wavelet power is used to examine modulation of one time series by another, or modulation of one frequency by another within the same time series. These figures are made by the average of the wavelet power spectra over all scales between the selected bands and they show a measure of the average year variance *vs* time.

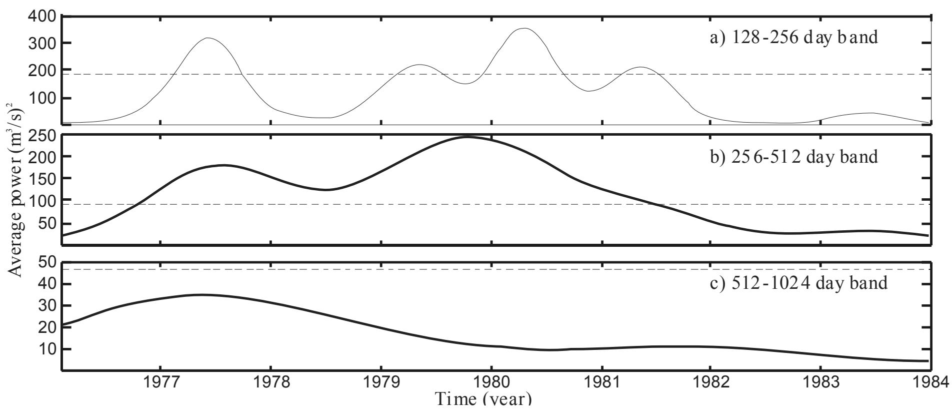

Fig. 4 Scale-average wavelet power for Piancó in Paraíba State over the (a) 128–256-day, (b) 256–512-day and (c) 512–1024-day bands. The dashed lines are the 95% confidence level assuming red-noise $\alpha = 0.75$.

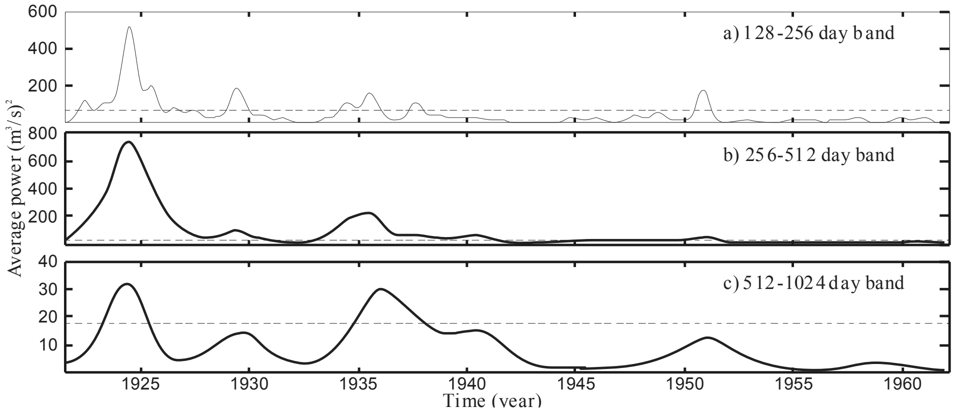

Fig. 5 Scale-average wavelet power for Serra Negra do Norte in Rio Grande do Norte State over the (a) 128–256-day, (b) 256–512-day and (c) 512–1024-day bands. The dashed lines are the 95% confidence level assuming red-noise $\alpha = 0.76$.

The distribution in time of the average variance of the signal can show which events are responsible for the peaks revealed in the global wavelet power spectra. For the studied basins, the results are similar to those of Figs 4 and 5, which depict the cases of Piancó (Fig. 4) and Serra Negra (Fig. 5), where the number of events with significant variance reduces with the reduction of the frequency. For Piancó, no event had significant variance in the biennial frequency. In a forecasting study, different models could be used for adjusting those events/periods corresponding to different frequency range.

DISCUSSION

This region is characterized by a semiarid climate, with the rainy season concentrated between February and May with the usual passage of the Inter-tropical Convergence Zone (ITCZ). Thus, it is natural that the semiannual frequency (periodicity at 182 days), annual frequency (periodicity at 365 days) present high values. However, this average situation can be disturbed a few times per-decade with the occurrence of the El-Niño phenomena and possible associated sea surface anomalies in the tropical Atlantic ocean. During these years, the precipitation regime in this hydrological basin (and throughout most of northeastern Brazil) can be highly perturbed with significantly less precipitation. These situations are responsible for high values of power spectrum in the semidecadal frequency (periodicity at 1825 days). Trends and decadal frequencies (periodicity at 3650 days) of this time series are confirmed by a time-average of power over time (Fig. 3), which shows the significant peaks above the 95% confidence level for the global wavelet spectrum, assuming red-noise, represented by the dashed lines. Although, the series show a decadal peak above the 95% confidence level, it corresponds to the power concentration within the cross-hatched region, where zero padding has reduced the variance.

All the runoff time series studied showed annual frequencies, but different to the rainfall time series, some of them presented semiannual and/or semidecadal frequencies, and these patterns can be used to determine the similar hydrological zones within the basin.

CONCLUSIONS

In order to study the variability of the total monthly rainfall and daily runoff time series in Piranhas-Açu River basin in northeastern Brazil, wavelet analysis was applied. The rainfall wavelet power spectra showed big power concentrations between the 8- to 16-day bands for all the sub-basins studied, as exemplified by the series of Angicos in Fig. 2, revealing an annual periodicity of such events. However, some breaks are observed in some year intervals (1925–1935 and 1965–1975), and semidecadal and 32-year frequencies are identified for the periods of 1945–1955 and 1940–1960, respectively. Periods with low variance in the 8- to 16-month band were identified, which are coincident with one of the major droughty events in that semiarid region of Brazil.

The modulation in separated periodicity bands (semiannual, annual and biennial bands) were done in order to extract additional information; e.g. the annual band was examined by an average of all scales between 256 and 512 days, giving a measure of the average daily variance *vs* time, where the periods with low or high variance could be identified. These time series could also be used instead of the original series in a forecasting study, where different models could be used for adjusting events corresponding to the selected time series. Finally, the main frequency components in the stream flow time series were studied with the global wavelet spectra, revealing how the stream flow frequency of each river is composed. Rivers within the same basin can present different stream flow characteristics, not easily detectable due to the similarity of the hydrological regime. Thus, the application of wavelet analysis could reveal those differences and help to identify significant frequency signals within the series, which is useful to determine the hydrological zones within a region.

Acknowledgments The authors are grateful to Dr Christopher Torrence of the Advanced Study Program at the National Center for Atmospheric Research, Colorado, USA and to the Brazilian agencies, ANA and ANEEL, for providing the wavelet analysis computer program and the hydrological data, respectively. This research had financial support from National Council for Technological and Scientific Development (CNPq–Brazil) and the second author is also a PIBIC research scholar of CNPq/UFPB.

REFERENCES

Farge, M. (1992) Wavelet transforms and their applications to turbulence. *Ann. Rev. Fluid Mech.*, **24**, 395–457.

Gaucherel, C. (2002) Use of wavelet transform for temporal characterisation of remote watersheds. *J. Hydrol.*, **269**(3–4), 101–121.

Graps, A. (1995) An introduction to wavelets. *IEEE Comp. Sci. Engng* **2**(2), 50–61.

Grossman, A. & Morlet, J. (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* **15**, 723–736.

Labat, D., Ronchail, J., Callede, J., Guyot, J. L., de Oliveira, E. & Guimaraes, W. (2004) Wavelet analysis of Amazon hydrological regime variability. *Geophys. Res. Lett.*, **31**(2).

Santos, C. A. G., Galvão, C. O., Suzuki, K. & Trigo, R. M. (2001) Matsuyama city rainfall data analysis using wavelet transform. *Ann. J. Hydraul. Engng*, *JSCE* **45**, 211–216.

Smith, L. C., Turcitte, D. L. & Isacks, B. L. (1998) Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrol. Processes* **12**, 233–249.

Torrence, C. & Compo, G. P. (1998) A practical guide to wavelet analysis. *Bull. Am. Met. Soc.* **79**(1), 61–78.