International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:6, No:7, 2012

Analysis of Precipitation Time Series of Urban Centers of Northeastern Brazil using Wavelet Transform

Celso A. G. Santos, Paula K. M. M. Freire

***Abstract*—**The urban centers within northeastern Brazil are mainly influenced by the intense rainfalls, which can occur after long periods of drought, when flood events can be observed during such events. Thus, this paper aims to study the rainfall frequencies in such region through the wavelet transform. An application of wavelet analysis is done with long time series of the total monthly rainfall amount at the capital cities of northeastern Brazil. The main frequency components in the time series are studied by the global wavelet spectrum and the modulation in separated periodicity bands were done in order to extract additional information, e.g., the 8 and 16 months band was examined by an average of all scales, giving a measure of the average annual variance versus time, where the periods with low or high variance could be identified. The important increases were identified in the average variance for some periods, e.g. 1947 to 1952 at Teresina city, which can be considered as high wet periods. Although, the precipitation in those sites showed similar global wavelet spectra, the wavelet spectra revealed particular features. This study can be considered an important tool for time series analysis, which can help the studies concerning flood control, mainly when they are applied together with rainfall-runoff simulations.

***Keywords*—**rainfall data, urban center, wavelet transform.

I. Introduction

THE wavelet transform is a recent advance in signal processing that has attracted much attention since its theoretical development in 1984 by [1]. 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 irregular 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], [6], [7].

C. A. G. Santos is with the Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil (phone: +55-83-3216-7684; fax: +55-83-3216-7179; e-mail: celso@ct.ufpb.br).

P. K. M. M. Freire is a former Master student of the Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil (e-mail: pulkymm@gmail.com).

The following sections describe the wavelet transform, the rainfall data of capital cities of northeastern Brazil, and then the application of wavelet to such data using the program developed by [3].

II. 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 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 the Morlet wavelet. 1. Such a wave 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.

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.

This single number gives a measure of the projection of this wave packet on the data during the 1911–1921 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 is defined as the product of a complex exponential wave and a Gaussian envelope:

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

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:

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

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 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 algorithm [3]: (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 (4), or whatever is appropriate for the mother wavelet in use, compute the daughter wavelet at that scale:

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

Where the ^ 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 (5), inverse transform back to real space;

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

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 (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.

III. Rainfall Data

The Northeast Region of Brazil is composed of nine states: Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE) and Bahia (BA), and it represents 18.26% of the Brazilian territory (Fig. 1). It has a population of 53.6 million people, which represents 28% of the total number in the whole country. Most of the population lives in urban areas and about 15 million people live in its semiarid region (*sertão*). It is famous in Brazil for its hot weather, beautiful beaches, rich culture (unique folklore, music, cuisine and literature), Carnival and St. John's festivities.

The capital cities are Aracaju – SE, Fortaleza – CE, João Pessoa – PB, Maceió – AL, Natal – RN, Recife – PE, Salvador – BA, São Luís – MA, and Teresina – PI. The biggest cities are Salvador, Fortaleza and Recife, which are the regional metropolitan areas of the Northeast, all with a population above a million inhabitants and metropolitan areas above 3.5 million.

Because for most of the year the Northeast is out of reach of the Intertropical Convergence Zone, the easterly trade winds blow across the region, giving abundant rainfall to the coast but producing clear, dry conditions inland where the escarpment blocks moisture flow. This gives rise to four distinct regions, the *zona da mata* on the coast, the *agreste* on the escarpment, *sertão* beyond and the mid north. The total monthly rainfall series since 1911 to February 2012 are shown in Fig. 2. The missing data are recorded as -1, which are the large space gaps in the figures.

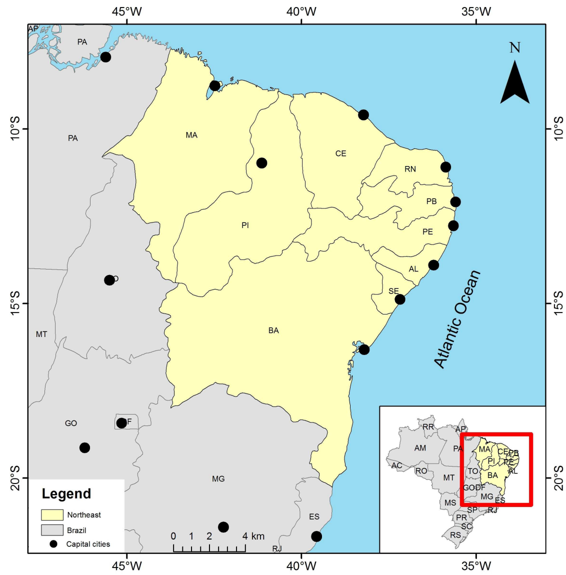

Fig. 1 Location of the capital cities of northeastern Brazil

IV. 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 fulfills 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 fulfill 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 unregularly 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.

A. Wavelet Power Spectrum

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

Fig. 3 shows the power (absolute value squared) of the wavelet transform for the monthly rainfall in the capital cities presented in Fig. 2, which is a record of more than 100 years. As stated before, 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 Fig. 3, it is clear that there is more concentration of power between the 8–16-month band, which shows that these time series have a strong annual signal. One should observe that the large empty spaces in Fig. 3 are cause by the missing data gaps (e.g. Figs 3e, 3f, 3g, 3h). The variance of power in 8–16-month bands (also confirmed later by Fig. 5) 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 in the beginning of 1980’s at João Pessoa city and a wet period from 1947 to 1952 at Teresina city.

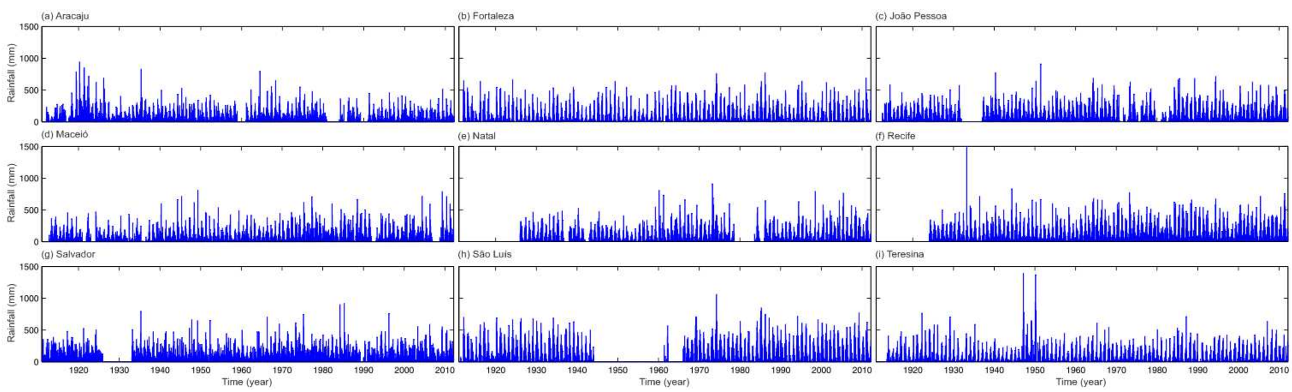

Fig. 2 Total monthly rainfall series from the capital cities of northeastern Brazil.

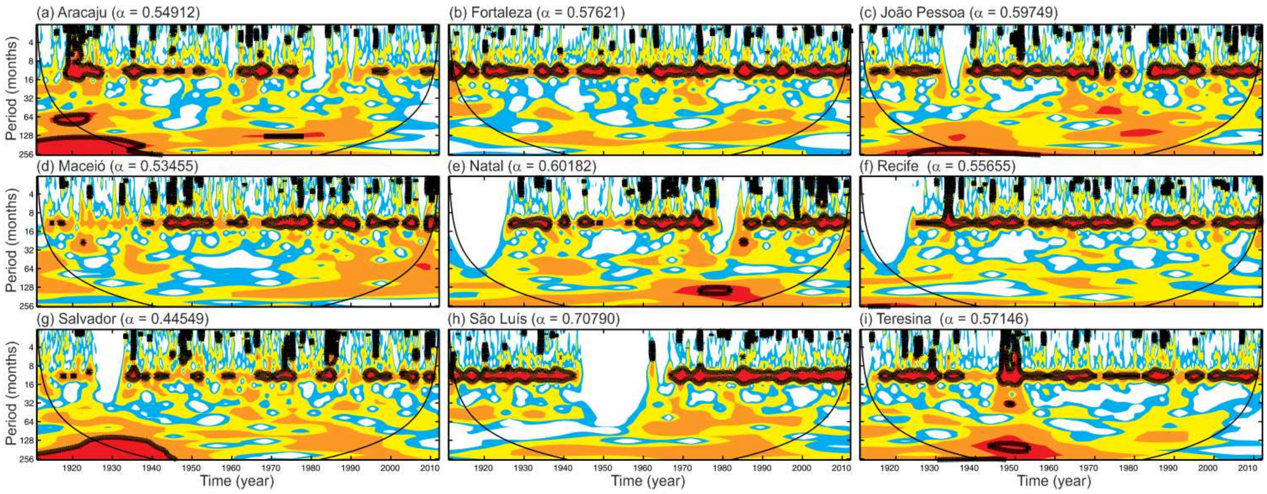

Fig. 3 The wavelet power spectra for each capital cities of northeastern Brazil. The contour levels are chosen so that 75%, 50%, 25%, and 5% of the wavelet power is above each level, respectively. Regions below the arcs are the cones of influence, where zero padding has reduced the variance. Black contour is the 5% significance level, using a red-noise (α) background spectrum. Wavelet power decreases according to the following order: red, orange, yellow, blue and white

The regions below the arcs in these figures are 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 arc 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 figures is the 5% significance level, using a red-noise background spectrum. Many geophysical time series can be modeled 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. 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.

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.

B. Global Wavelet Power Spectrum

The annual frequency (periodicity at 12 months) of these time series are confirmed by an integration of power over time, which show only one significant peak above the 95% confidence level for the global wavelet spectra, assuming red-noise, represented by the dashed lines (Fig. 4). However, Fig. 4 also presents almost significant peaks (at the 5% level) centered in the 2–4-month band. In fact, most extreme monthly precipitation values for those cities (values above 250 mm in Fig. 2) correspond to pulses of highly significant power within the 2–4-month band (Fig. 3b). 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 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. For instance, despite the difference among the observed hyetogaphs, the similarity of the global wavelet spectra shows that these series belong to the same region (Northeast).

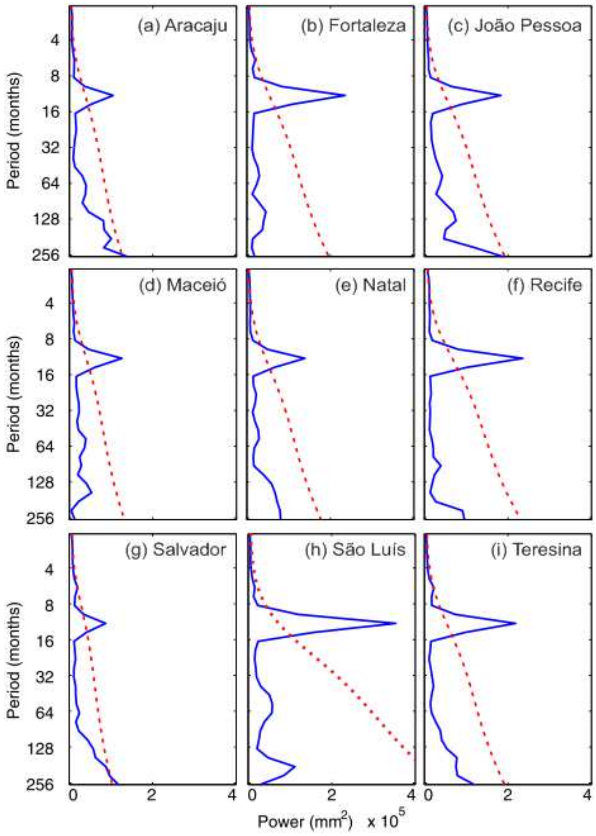

Fig. 4 The global wavelet power spectra (blue line). The red dashed line is the 5% significance level for the global wavelet spectra

C. Scale-average Time Series

The scale-average wavelet power (Fig. 5) 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. These figures are made by the average of Fig. 3 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., a dry period can be identified in the 1950's at Fortaleza city and a wet period since 2000 at São Luís city. A dendrogram for those data is presented in Fig. 6.

V. Conclusion

In order to study the variability of the monthly rainfall time series in the capital cities of northeastern Brazil, wavelet analysis was applied. The wavelet power spectra show 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 high variance in such a band could 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. The wavelet power spectra showed that Maceió and Salvador cities have similar rainfall patterns, Aracaju shows some similarity as well, whereas João Pessoa, Natal and Recife cities form another group. Fortaleza, Teresina and São Luís have unique characteristics, and then they cannot be included in or form another group, although Fortaleza is the closest one to the João Pessoa's group.

Finally, further study could include stream flow analysis in order to benefit runoff-erosion models [8] from a quantitative breakdown of the temporal components of stream flow.

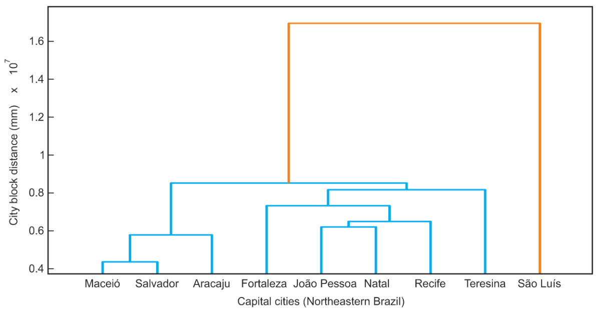

Fig. 6 Dendrogram for the scale-average wavelet power 8–16-month band of northeastern Brazilian capital cities

Acknowledgment

The writers wish to thank Dr. Christopher Torrence of Exelis Visual Information Solutions, Colorado, USA, for providing the wavelet analysis computer program and material, as well as for his comments. The authors also thank Brazilian National Water Agency (ANA) for the precipitation data and to CNPq for the financial support.

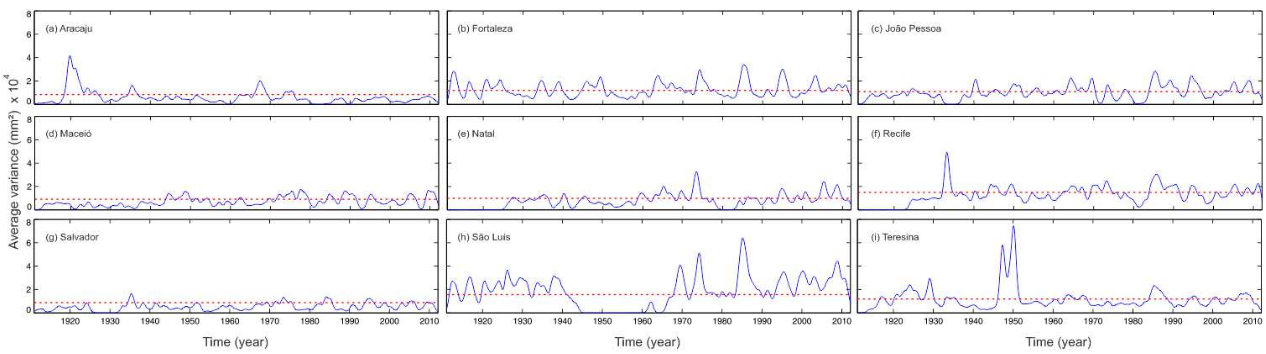

Fig. 5 Scale-average wavelet power over the 8–16-month band for the total monthly rainfall in each capital city.The dashed line is the 95% confidence level assuming the respective red noise

References

[1] A. Grossman, and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, pp. 723–736, 1984.

[2] A. Graps, "An introduction to wavelets," *IEEE Computational Science and Engineering*, vol. 2, No. 2, pp. 50–61, 1995.

[3] C. Torrence, and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Amer. Meteor. Soc.*, vol. 79, No. 1, pp. 61–78, 1998.

[4] M. Farge, "Wavelet transforms and their applications to turbulence," *Ann. Rev. Fluid Mech.*, vol. 24, pp. 395–457, 1992.

[5] L. C. Smith, D. L. Turcitte and B. L. Isacks, "Stream flow characterization and feature detection using a discrete wavelet transform," *Hydrological Processes*, vol. 12, pp. 233–249, 1998.

[6] I. Y. L. G. Braga, and C. A. G. Santos. "Viability of rainwater use in condominiums based on the precipitation frequency for reservoir sizing analysis," *J. Urban and Environ. Engng*, vol. 4, No. 1, pp. 23–28, 2010.

[7] C. A. G. Santos, B. S. Morais, and G. B. L. Silva. "Drought forecast using Artificial Neural Network for three hydrological zones in San Francisco river basin," *IAHS Publication*, vol. 333, pp. 302–312, 2009.

[8] C. A. G. Santos, V. S. Srinivasan, K. Suzuki, and M. Watanabe, "Application of an optimization technique to a physically based erosion model," *Hydrological Processes*, v. 17, No. 5, pp. 989–1003, 2003.

Celso A. G. Santos was born in Campina Grande, Paraíba, Brazil. He is graduated in Civil Engineering, 1990, and Data Processing, 1991, from the Federal University of Paraíba, Brazil. He got his master degree in Civil and Environmental Engineering, 1994, and doctor in Engineering, 1997, both from Ehime University, Ehime, Japan. His major field of study is hydrology.

He was a postdoctoral fellow of Japan Society from Promotion of Science (JSPS) from 1997 to 1999. He was an associate professor at Ehime University from 1999 to 2001. Since 2002, he is an associate professor at Federal University of Paraíba. His research interest includes runoff-erosion modeling, model optimization using technique such as genetic algorithm, swarm particles, simulated annealing and differential evolution optimizer, wavelet transform, and artificial neural networks.

Dr. Santos is member of the Brazilian Association of Water Resources, and the International Association of Hydrological Sciences, and has published several papers at IAHS Publication, Hydrological Processes, Natural Hazards, Water Science and Technology, J. Urban and Environmental Engineering, J. Hydroscience and Hydraulic Engineering, J. Hydraulic, Coastal and Environmental Engineering, and Brazilian J. Water Resources, for example.

Paula K. M. M. Freire was born in João Pessoa, Paraíba, Brazil. She is graduated in Internet Systems, 2009, from the Federal Institute of Paraíba, and in Civil Engineering, 2011, from the Federal University of Paraíba. She got her master degree in Urban and Environmental Engineering, 2012, from Federal University of Paraíba.

She is a research fellow of Brazilian Council for Scientific and Technological Development (CNPq) since 2009 at the Federal University of Paraíba. Her major field of study is model optimization and hydrology.

Freire is member of the International Association of Hydrological Sciences, and has published papers at IAHS Publication, Water Science and Technology, Journal of Urban and Environmental Engineering, and Land Reclamation.