ICIWRM – 2000, Proceedings of International Conference on Integrated Water Resources
Management for Sustainable Development, 19 – 21 December, 2000, New Delhi, India

Theme Paper *Theme V*

Periodicities in annual hydrologic data from the midwestern USA

A. RAMACHANDRA RAO
School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA
KHALED HAMED
Irrigation and Hydraulics Department, Faculty of Engineering, Cairo University, Giza, Egypt

Abstract

Many investigators have reported the existence of periodicities, other than the annual and semi-annual, in hydrologic and climatic data. These results are based on the spectral analyses of data. Spectral analysis has several well known drawbacks, to overcome which the multitaper method (MTM) of spectral analysis was developed. The MTM has the additional advantage that the spectral components can be tested for statistical significance. In the present study, the MTM is used to investigate the periodicities in the annual streamflow, rainfall, temperature and tree ring data in the midwestern United States. The results indicate the existence of weak periodicities of periods less than 50 years.

INTRODUCTION

Conventionally, hydrologic processes have been regarded as continuous spectrum processes such as the autoregressive (AR), autoregressive-moving average (ARMA), and other models (Salas et al., 1988). No line components have been recognized in hydrologic data except for the well established, physically based daily and monthly cycles. However, several studies suggest that other cycles with different frequencies may exist in hydrologic as well as temperature data (Currie, 1990; Rao and Yu, 1989; Labitzke and Van Loon, 1990; Mitra et al., 1991; Burroughs, 1992; Mitra and Dutta, 1992; and Lall and Mann, 1995). These cycles may represent the influence of a number of natural phenomena such as the sunspot cycle, the sun's magnetic (Hale) cycle, El-Niño/Southern Oscillations (ENSO), Quasi-Biennial Oscillations, and others. However, the analysis methods used in previous studies do not offer a measure of statistical significance for different cyclic components. The amplitude of spectral line components is usually the only criterion for identifying such line components, giving the chance that spurious peaks may be mistaken for true cycles. In contrast, the multitaper method (MTM) offers a statistical test for assessing the significance of different components and provides a means for distinguishing between spurious and actual line components in the estimated spectrum, independent of their amplitude.

In this paper, riverflow, rainfall, and temperature data as well as tree-ring data from different locations in the Midwestern United States are analyzed. The MTM is used in this analysis to identify periodic components in these time series and assess their statistical significance. Results from different time series are discussed and compared with those from previous studies, and the relation between results from different data sets are investigated.

MULTI-TAPER METHOD (MTM) OF SPECTRAL ANALYSIS

The MTM method (Thomson, 1982) makes use of an extended version of the spectral representation in Eq. 1.

$$x_t = \int_{-1/2}^{1/2} e^{i\,2\pi f\,t}\,\mathrm{dz(t)} \tag{1}$$

The process x_t in this case may include a number of periodic components in addition to the underlying stationary process as in Eq. 2 (Percival and Walden, 1993),

$$\begin{aligned} x_t &= \sum_j C_j \cos(2\pi f_j t + \phi_j) + \zeta_t \\ &= \sum_j \mu_j\, e^{i2\pi f_j t} + \mu_j^{*}\, e^{-i2\pi f_j t} + \zeta_t \end{aligned} \tag{2}$$

where ζ_t (the background continuum) is a zero mean stationary process with spectral density function $S(f)$ that is not necessarily constant, i.e., it can be colored noise, C_j and f_j are the amplitude and frequency of periodic or line component j, $\mu_j = (C_j/2)\ e^{i\phi_j}$ is the complex amplitude corresponding to the real amplitude C_j, and (*) indicates complex conjugate. These types of processes are known as centered stationary or conditionally stationary processes and are often noted as having mixed spectra (Thomson, 1990a). The details of the MTM are found in Thopmson (1982, 1990a, b).

Data Used in the Study

Data used in this study are divided into four categories: flow, rainfall, temperature, and tree-ring data. Flow data are obtained from the United States Geological Survey (USGS) database. Both rainfall and temperature data are obtained as two different sets. The first set is obtained from the Historical Climatology Network (HCN) database for individual stations. The second set is obtained from the National Climatic Data Center (NCDC) database or state divisions. Tree-ring data are obtained from the National Oceanic and Atmospheric Administration (NOAA) database.

Flow time series from 26 stations in the study region were obtained from the USGS database. These consist of 6 stations in Iowa, 7 stations in Illinois, 6 stations in Indiana, and 7 stations in Ohio.

Rainfall and temperature time series from 36 stations were obtained from the HCN database. These consist of 9 stations in Iowa, 10 stations in Illinois, 8 stations in Indiana, and 9 stations in Ohio.

NCDC rainfall and temperature data are available as divisional averages. Each state is divided into a number of divisions as follows: 9 divisions in Iowa, 9 divisions in Illinois, 9 divisions in Indiana, and 10 divisions in Ohio. The divisional average data are obtained by averaging the records from a number of stations within that division. NCDC Divisional average data are available for all divisions from 1895 to 1992 for rainfall data, and from 1895 to 1995 for temperature data.

For treeing data, 22 time series were chosen from the database. These consist of 9 series in Iowa, 9 series in Illinois, one series in Indiana, and 3 series in Ohio. All the series are annual series and are at least 100 years long.

ANALYSIS OF HYDROLOGIC AND CLIMATIC DATA

For the sake of consistency, the number of frequency points used for all data is set at 2048, which is sufficient to prevent circular correlation and ensure accurate results of the F-test for the significance of peaks in spectra.

Results From MTM Analysis

When a large number of regional data are available, additional inference can be made about periodicities. Generally speaking, a periodicity which is found to be significant in a large number of stations in a given region suggests that such a component is evident as a regional periodicity. In statistical terms, if the number of stations with such periodicity is significantly larger than the expected number under the null hypothesis of no periodicity, then this periodicity is statistically significant; otherwise it is spurious. The expected number of stations under the null hypothesis is simply equal to the significance level α times total number of stations. A 95% confidence interval can then be constructed for the average number by considering the binomial distribution with a rate of failure of α.

The use of a significance level of $1/N$ recommended by Thomson (1990a) seems to be too strict for such an analysis, since the goal is not to assess significance for a single station but rather to compare the actual and expected number of occurrences of significant components in the studied region. Therefore a statistical significance level α of 0.1 (probability of obtained F-value is 0.90 or more) was considered for the significance of line components in addition to the $1/N$ level (probability equal to or exceeding $1-1/N$), where N is the number of data points. Figures 1 and 2 are typical of the results obtained. Components which are significant at the 0.1 level in Figs. 1 and 2 are shown as circles which are scaled between 0.9 and 1.0 according to the probability of their F-test. Components which are significant at the $1/N$ level are further identified by a plus sign within their respective circles. The following results were obtained from MTM analysis:

a) USGS Flow Data: The results from the MTM analysis of the USGS flow data are shown in Fig. 1 for band width values of $2.5/N$ and $4/N$, respectively. Periodicities found in the USGS flow data are summarized in Table 1. The first column in Table 1 gives the approximate range of each group of periodicities. Columns 2 to 5 pertain to the case of $W=2.5/N$ and give the number of stations with significant components at the 0.1 level, the number of stations with significant components at the $1/N$ level, the mean of the period in the respective group, and the average percentage of the total variance explained by components in that group. Columns 6 to 9 give the same information but with regard to the case of *W*=4/*N*. For this group of 26 stations, considering a 95% confidence interval, no more than five stations should have significant components at a given frequency at the 0.1 level under the null hypothesis of no periodicity. Furthermore, no more than one station should have significant components at a given frequency at the 1/*N* level under the null hypothesis of no periodicity. Based on the numbers in Table 1, sufficient evidence of periodicities does not exist in the range from 89 to 121 years, although a component with such periodicity would explain 8 to 15% of the total variance in about seven stations. Most of these stations are found in Iowa, which raises the question of a possible spatial variation of periodicities. The group of 24 to 30 years with an average period of 25.7 years is found in more stations than would be only due to chance. Periodicities of 10-12, 7-8, and 4-5 also seem to be evident. Periodicities of 2-3 years seem to be less evident at the 1/*N* level, with the groups centered at 2.1 and 2.6 being the most common at the 0.1 level. The percentage of total variance explained by the identified periodicities range from 4 to 14%. The largest contribution comes from the 5.7 years component at 13-14%, followed by the 11.1 years component at about 12% of the total variance.

b) HCN Rainfall Data: The results from the MTM analysis of the HCN rainfall data were analyzed similar to the river flow data. For this group of 36 stations, considering a 95% confidence interval, no more than seven stations should have significant components at a given frequency at the 0.1 level under the null hypothesis of no periodicity. Furthermore, no more than one station should have significant components at a given frequency at the 1/*N* level under the null hypothesis of no periodicity. Based on these results there is little evidence of periodicities in the range of 100-150, 30-40, 20-25 years. Periodicities of 10-12, 5-6, 4-5, and 2-4 are the most common in terms of the number of stations having significant components at both the 0.1 and 1/*N* levels. The percentage of total variance explained by these periodicities range from 5 to 15% with a maximum of 15-16% for the group of 100-150 years followed by the 30-40 years group at 7-8% of the total variance.

c) NCDC Divisional Average Rainfall Data: For this group of 37 divisions, considering a 95% confidence interval, no more than seven divisions should have significant components at a given frequency at the 0.1 level under the null hypothesis of no periodicity. Furthermore, no more than one division should have significant components at a given frequency at the 1/*N* level under the null hypothesis of no periodicity. Based on the results of these data the number of stations with periodicities of 90 to 140 years is too small to support the existence of such periods in the data. The same is more or less true with periods of 30-40, 20-25, and 17-20 years. The most significant groups are 2.6-2.8, 3-3.5, 4-5, and 5.4-6 years. The percentage of total variance explained by these periodicities range from 5 to 8%, with the largest contribution from components 5.4-6 years at 7 to 8% of the total variance.

It was observed that there is a switch of periodicity from around 5.8 years in Ohio, Indiana, and most of Illinois to 6.4 years in Iowa and the first four divisions in Illinois. This, again, raises the question of spatial variability of periodicities. Another example is periods of 30 to 33 years in which seem to be found only in Indiana and Ohio and almost disappear in Iowa and Illinois in exchange for periodicities of 19 to 21 years.

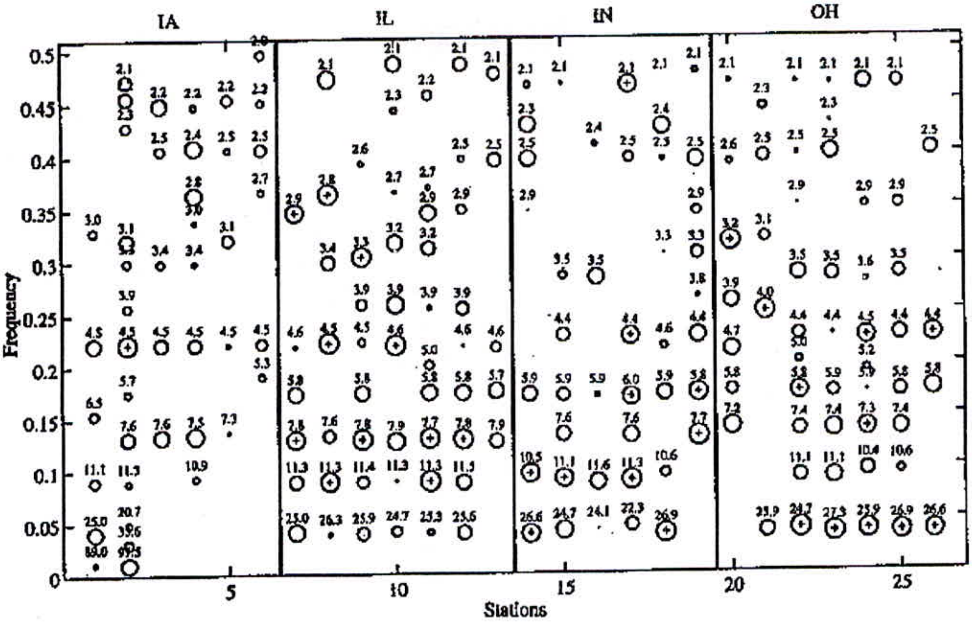

Figure 1. Periodicities in the USGS Flow Data Using MTM Analysis with Band Width W=2.5/N.

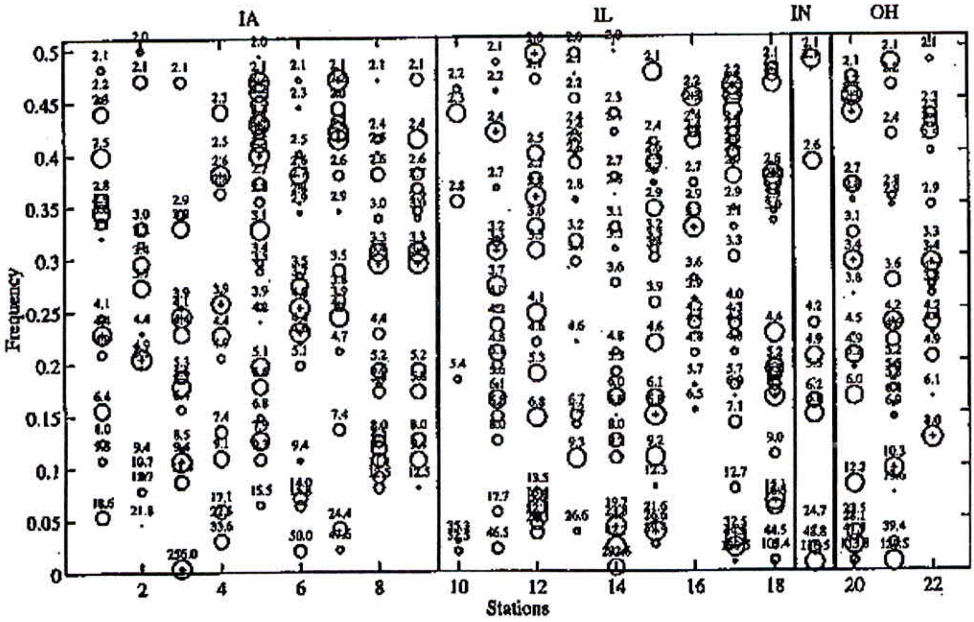

Figure 2. Periodicities in the NOAA Treering Data Using MTM Analysis With Band Width W=2.5/N.

Table 1. Periodicities in the USGS Flow Data, MTM Analysis.

Periodicity	W = 2.5/N				W = 4/N			
Group (years)	n[1] at 0.1	n[1] at 1/N	Mean Period	% Var.	n[1] at 0.10	n[1] at 1/N	Mean Period	% Var.
89-121	2	0	93.3"6.0	8	7	0	109.2"8.5	15
24-30	17	7	25.7"1.0	10	13	8	27.3"1.2	9
10-12	18	5	11.1"0.3	12	18	1	10.9"0.4	12
7-8	19	6	7.6"0.2	10	4	2	7.8"0.1	13
5-6	21	3	5.7"0.2	13	14	4	5.7"0.2	14
4-5	24	6	4.5"0.2	9	14	7	4.5"0.2	10
3.5-4	13	1	3.7"0.2	4	12	1	3.7"0.2	5
3-3.5	14	2	3.2"0.1	5	16	4	3.3"0.1	5
2.8-3	9	2	2.9"0.03	5	10	0	2.9"0.05	6
2.4-2.8	22	0	2.6"0.1	4	22	3	2.6"0.09	4
2.25-2.4	6	0	2.3"0.04	4	6	0	2.3"0.03	4
2.15-2.25	8	0	2.2"0.03	4	11	1	2.2"0.03	4
2-2.15	16	1	2.1"0.04	4	20	4	2.1"0.03	4

[1] *n* is the number of stations with components that are significant in a given periodicity group at the specified significance level.

d) HCN Temperature Data:

Based on the results of these data, there is too little evidence to establish periods of 90 to 120 years. The most significant groups are 50-80, 17-20, 10-12, 5-6, and 2-2.2 years. Other periods seem to be much more significant for the case of *W*=2.5/*N* than for the case of *W*=4/*N*. This may be attributed to a narrower bandwidth of the temperature data for which *W*=4/*N* is large enough to smooth these periodicities. The percentage of the total variance explained by these periodicities ranges from 3 to 14%. The largest contribution comes from periods of 90-120 years at 14%, followed by periods of 50-80 years at 12-14%, then 10-12 years at 11% of the total variance.

e) NCDC Divisional Average Temperature Data:

Periods from 50 to 80 years are strongly significant for *W*=2.5/*N* at both the 0.1 and 1/*N* levels. For *W*=4/*N* only one station is significant at the 1/*N* level. Again, this may be attributed to the narrowness of the actual band width of periodicities in the temperature data. Periods of 10-12 years and 2-2.15 years are also highly significant. Other common periods are 2.2-2.4, 2.9-3.2, and 5-6 years. The percentage of total variance explained by these periodicities range from 2 to 14%. The largest contribution is from periods of 50 to 80 years at 12-14%, followed by periods 10 to 12 years at 9-11%, then 2.4 to 2.6 years at 8-9% of the total variance.

As was the case with the NCDC divisional rainfall data, some periodicities seem to vary spatially. Periodicities of 18 to 19 years seem to disappear in Iowa and the first two divisions in Illinois. Also, there seems to be an east-to-west drift in the 10 to 11 years periodicities from 10.6 in Ohio to 10.9 in Iowa. The same is observed for periods of 6 to 7 years which consistently change from 6.5 to 6.7 from east to west.

f) NOAA Tree-ring Data:

Since the length of the times series in this group is more than 200 years, a significance level of 0.95 instead of 0.90 was considered as well as the 1/*N* significance level. In these data, groups of periodicities can hardly be identified (Fig. 2). Also, very few of the observed periodicities are significant at the $1/N$ level. The MTM analysis was repeated for the last 100 years of the NOAA tree-ring data. A much clearer grouping of periodicities can be seen in these figures, and periodicities that are significant at the $1/N$ level are more common than in the case when the whole record is analyzed.

In general, one would expect periodicities to be more well-defined as the length of the records increases. The fact that periodicities were not well grouped when the entire record is considered for MTM analysis is counterintuitive. One reason for such behavior may have to do with the quality of the earlier part of the tree-ring data. Another concern is whether a different type of nonstationarity is involved, in which not only periodic components are added to a stationary process, but also the amplitude and frequency of such components change with time.

Groups of periodicities in the last 100 years of the NOAA tree-ring data were analyzed. The most common periods were 2.5, 3.3, 3.8, and 10-12 years. The percentage of total variance explained by these periodicities range from 5 to 18%. Percentages of the total number of stations with significant periodicities are given in Table 2.

CONCLUSIONS

In light of the results discussed above, the following conclusions are presented.

It is evident that significant components of different periods are present in the data. As such, these processes cannot be considered stationary processes with continuous spectra, but rather centered stationary processes with mixed spectra.

The most common four periodicities in all data are 2.5 to 2.6 years, 3 to 3.5 years, 5 to 6 years, and 10.7 to 11.1 years. It was also found that periodicities of 17.9 to 19.6 years and 60 to 69 years are found mostly in temperature data, while periodicities of 21 to 26 years are found mostly in flow data, and periodicities of 32 to 34 years are found mostly in rainfall data.

The percentage of the total variance explained by different periodicities varies from 3 to 18%. High frequency components contribute 3 to 6% of the total variability, while there seems to be an increase of percentage for lower frequency components up to 18% of the total variability.

It was observed in some data sets, especially divisional average data, that some type of frequency drift occurs in the study region from east to west. In some cases frequencies are discontinued and other frequencies evolve also moving from east to west. This wouldsuggest some type of spatial distribution of frequencies. However, a much larger database would be required to further investigate this behavior.

Based on the results of the harmonic analysis of tree-ring data, there is an indication of a different type of nonstationarity. This type of nonstationarity involves changes in the structure of the process where both the frequency and amplitude of cyclic components are time dependent. This type of nonstationarity needs further investigation through time-frequency analysis.

Table 2. Summary of the Periodicities in the Data.

The Percentages of the Total Number of Stations with Significant Periodicities are given.

Periods	USGS Flow	Rainfall		Temperature		Tree-ring
		HCN	DIV	HCN	DIV	
2.1	77%	39%	43%	86%	86%	68%
2.2	42%	39%	62%	---	---	---
2.3	23%	67%	78%	78%	100%	59%
2.5-2.6	100%	36%	41%	72%	97%	68%
2.7-3	38%	64%	65%	---	32%	45%
3-3.5	62%	86%	78%	92%	97%	82%
3.5-4	50%	83%	46%	47%	86%	100%
4-5	92%	92%	78%	31%	68%	32%
5-6	81%	72%	68%	83%	95%	45%
6-7	---	28%	30%	86%	95%	32%
7-9	73%	50%	46%	83%	89%	36%
10.7-11.1	69%	47%	57%	97%	100%	36%
17.9-19.6	---	---	---	61%	65%	---
21-26	65%	31%	24%	---	---	73%
32-34	---	22%	46%	---	---	---
65-69	---	33%	---	78%	100%	---
89-150	27%	33%	22%	31%	----	---

References

Burroughs, W.J., 1992. Weather Cycles: Real or Imaginary. Cambridge University Press.

Currie, R.G., and O'Brien, D.P., 1990. "Deterministic signals in precipitation records from American Corn Belt." International Journal of Climatology, Vol. 10, pp. 179-189.

Labitzke, K., and VanLoon, H., 1990. "Association between the 11-year Solar cycle, the Quasi-Biennial Oscillation, and the atmosphere: A summary of recent work." Phil. Trans. Royal Society Lond., A330, pp. 577-589.

Lall, U., and Mann, M., 1995. "The Great Salt Lake: A barometer of low-frequency climatic variability." Water Resources Research, 31(10), pp. 2503-2515.

Mitra K., and Dutta, S.N., 1992. "18.6-year Luni-Solar nodal an 10-11-year Solar signals in rainfall in India." Int. J. Of Climatology, 12, pp. 839-851.

Mitra, K., Mukherji, S., and Dutta, S.N., 1991. "Some indication of 18.6 year Luni-Solar and 10-11 year Solar cycles in rainfall in North-West India, the plains of Uttar Pradesh and North-Central India." Int. J. Of Climatology, 11, 645-652.

Percival, D.B., and Walden, A.T., 1993. "Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques". Cambridge University Press, Cambridge, U.K.

Rao., A.R., and Yu, G.H., 1989. "Hale cycle and Indian drought and flood area indexes." ASCE J. Irr. And Drainage Eng., 115(3), pp. 315-333.

Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L., 1988. Applied Modeling of Hydrologic Time series. Water Resources Publications, Littleton, CO.

Thomson, D.J., 1982. "Spectrum estimation and harmonic analysis." Proc. IEEE, 70, pp. 1055-1096.

Thomson, D.J., 1990a. "Time series analysis of Holocene climate data." Phil. Trans. Roy. Soc. London, Series A, 330, pp. 601-616.

Thomson, D.J., 1990b. "Quadratic-inverse spectrum estimates; Applications to paleoclimatology." Phil. Trans. Roy. Soc. London, Series A, 332, pp. 593-597.