Mausam, (1990), 41, 2, 273-278

551.501.89 : 551.553.21

Intraseasonal monsoon fluctuations seen through 25 years of Indian radiosonde observations

JON AHLQUIST, VIKRAM MEHTA, ANDREW DEVANAS and THOMAS CONDO

Dept. of Met., Florida State Univ.,

Tallahassee, Florida 32306-3034, U. S. A.

सार – मानसून के अन्तराऋतुनिष्ठ क्रिया के लिए 12 मुख्य भारतीय नगरों के रेडियोसौन्दे रिकार्डों का गुणता की जांच व विश्लेषण किया गया है। प्रतिदिन दो प्रेक्षणों से काल श्रेणी बना कर निकृष्ट मान को हटाते हुए और रिक्त आंकड़ों को रेखिक अन्तर्वेशन से भरते हुए छनन (फिल्टरिंग) के साथ और उसके बिना, काल श्रेणी को आलेखित किया गया है और अन्तराऋतुनिष्ठ उच्चावचन और उनकी अन्तर वार्षिक परिवर्तन शीलता की प्रकृति को प्रदर्शित करने के लिए स्पैक्ट्रल विश्लेषण किया गया है अधिकतम एंट्रापी प्रणाली (MEM) का उपयोग करते हुए काल श्रेणी से स्पेक्ट्रा का आकलन किया गया जो कि काल श्रेणी को स्वसमाश्रयी (AR) निदर्श समंजन करती है दशांश श्रेणी निदर्श पर आधारित स्पेक्ट्रा दिखाते हैं कि मानसून मौसम में अधिकतर अन्तर अन्तराऋतुनिष्ठ क्रिया से आते हैं जिनका काल माप 10 दिन से अधिक है किन्तु अधिकांश ग्रीष्म के लिए 10-20 एवं/या 30-35 दिन काल माप अलग चरम को नहीं दर्शाते हैं।

ABSTRACT. Radiosonde records for 12 major Indian cities have been checked and analyzed for intraseasonal activity in the monsoon. After forming time series with two observations per day, removing bad values, and filling data gaps with linear interpolation, the time series have been plotted with and without filtering and have been subjected to spectral analysis to reveal the nature of intraseasonal fluctuations and their interannual variability. Spectra were estimated from the time series using the maximum entropy method (MEM), which fits an autoregressive (AR) model to the time series. MEM spectra based on tenth order AR models show that most of the variance in monsoon weather comes from intraseasonal activity with periods longer than 10 days, but do not show separate peaks at 10-20 and/or 30-50 day time scales for the majority of summers.

1. Introduction and description of data set

Only a few studies, *e.g.*, those by Mehta and Ahlquist (1986) and Mehta and Krishnamurti (1988), have looked at intraseasonal fluctuations in the Indian summer monsoon using several years of data. The purpose of our investigation was to examine the importance of intraseasonal monsoon activity using the longest multilevel data set known to the authors.

The data chosen were Indian radiosonde records obtained on magnetic tape from the U.S. National Center for Atmospheric Research. The data on this tape end on 30 June 1978. Soundings for some cities, like New Delhi date back, with few gaps, to 1 January 1951, while a few cities have less than a hundred soundings. The records for 12 major Indian cities were chosen for study : Ahmedabad, Allahabad, Bombay, Calcutta, Gauhati, Jodhpur, Madras, Nagpur, New Delhi, Port Blair, Trivandrum and Visakhapatnam. Each of these cities radiosonde records is at least a decade long, most are over 15 years long, and a few are over 25 years long.

Twice daily time series were formed for five variable (zonal wind speed, meridional wind speed, temperature, geopotential height and dewpoint temperature) at eight levels (100, 85, 70, 50, 40, 30, 20 and 10 kPa—multiply by 10 convert kPa to mb). At this point in the procedure, any missing data in the time series were filled with a missing data indicator equal to —999 (which could never occur as a legitimate value for any of the variables).

Two quality control tests were imposed. The first required that all observations fall within an acceptable range of values. The acceptable range of values for each variable was determined subjectively as follows. A histogram was prepared and plotted for each variable which showed the distribution of time series values. From these histograms, it was possible to see, at a glance, what were reasonable minimum and maximum limits for legitimate data, while flagrantly small or large values stood out prominently. Limits, subjectively chosen based on the histograms, were used to screen the data, and any values which lay outside of the "acceptable" ranges were replaced by the missing data value —999.

This first test was not enough to catch all the bad data. We also looked for spikes in the time series. A "spike" was defined as a single datum whose value was markedly different (the precise amount was adjusted for each variable) from the value that preceded it and the value that followed it. Any spike was replaced by the missing data value —999. We have not performed any inter-level checks such as looking for super-adiabatic lapse rates.

At this point, the time series were scanned again numerically, and all the missing data values of —999 were replaced by linear interpolation in time.

Since intraseasonal activity was our primary interest, the mean seasonal cycle was removed from each of the time series. The mean seasonal cycle was defined as a constant *plus* sinusoids with periods of 1, 2, 3, and 4 cycles per year. The seasonal cycle was determined by least squares methods.

2. Maximum entropy method

2(a). *Spectral analysis*

Most previous studies of intraseasonal monsoon activity have reported 10-20 and 30-50 day activity. So far, our time series have not produced many satistically significant 10-20 and/or 30-50 day spectral peaks, so we want the reader to view our results with caution. Hence, we explain our spectral analysis methods in detail.

By definition, the spectrum of a random process is the Fourier transform of the process's autocovariance function (Jenkins and Watts 1968). One of the key difficulties in estimating the spectrum of an observed time series is that observations are of finite length, so that the autocovariance can be estimated for only a finite number of lags. Estimating the spectrum by Fourier transformation of the estimated autocovariance assumes that the autocovariance is zero for all lags larger than those at which the autocovariance is estimated. Alternatively, one may estimate a spectrum by smoothing the periodogram, but computation of the periodogram assumes that the data are periodic. Both of these methods involve unrealistic assumptions.

The Maximum Entropy Method (MEM) handles the problem of limited knowledge of the time series in a different manner. MEM, reviewed by Ulrych and Bishop (1975) and Papoulis (1981), extrapolates the autocorrelation to infinite lag using a statistical model which maximizes the "entropy" (uncertainty) of the extrapolated values while being consistent with the known, *i.e.*, estimated, autocorrelation values. As shown by van den Bos (1971), the same statistical model would be determined by least squares fitting an autoregressive (AR) model to the observed time series. AR modelling is a special case of autoregressive, integrated, moving average (ARIMA) modelling, which is covered in detail by Box and Jenkins (1976). The reader is also referred to Jaynes' (1982) discussion of the relation between maximum entropy and other methods of spectral analysis and to Bendat and Piersol's remarks (1986) regarding spectra based on AR, MA, and ARIMA models.

An AR process (x_t) of order N driven by white noise (n_t) is of the form :

$$x_t = \sum_{k=1}^{N} a_k \, x_{t-k} + n_t \tag{1}$$

where the subscript 't' indicates time. The coefficients $\{a_k\}$ for this model can be determined using subroutine YWPR* or MEMPR which appear in the appendix to Ulrych and Bishop (1975).

Without approximation, the spectrum of an AR model is:

$$S(f) = \frac{2\sigma_n^2 \, \triangle t}{\left| 1 - \sum_{k=1}^{N} a_k \exp\left(-i\, 2\pi k f \triangle t\right) \right|^2}$$

$$\text{for } 0 \leqslant f \leqslant f_N \tag{2}$$

where $f_N \equiv 1/(2\triangle t)$ is the Nyquist frequency and $\sigma_n{}^2$ is the variance of the white noise process $\{n_t\}$.

Traditionally, there have been two problems with MEM spectral estimation : (a) choosing the best order N for the AR model and (b) estimating spectral confidence limits. Ulrych and Bishop (1975) recommend the value of N which minimizes Akaike's Final Prediction Error (FPE); this selects the AR model which, if used for forecasting, would have the least mean squared error [*see* Ulrych and Bishop (1975) for details]. Less has been written about MEM spectral confidence limits, to which we now turn.

2(b). *Confidence limits for MEM spectra*

There is no guarantee that the spectrum of a random process must be smooth. Nonetheless, any time we estimate a spectrum by smoothing a periodogram or a transformed autocovariance, we are estimating as smoothed version of the spectrum [Jenkins and Watts 1968, (6.3.30), p. 243]. Likewise, if we use MEM and choose a model order that is too small, then our spectral estimate will be a smoothed version of the actual spectrum.

We must keep in mind that confidence limits on an estimated spectrum do not represent our confidence that the actual spectrum lies within the indicated bounds; rather, they represent our confidence that the spectrum *smoothed by the same procedure as used to stabilize our spectral estimates* lies within the indicated bounds. At first glance, one may think it best to choose a large order AR model for all MEM spectral estimates so that the estimated spectrum will not be smoothed too much. The trade-off for this increased resolution is a lack of stability in the spectral estimates, *i.e.*, possibly spurious peaks and wider confidence limits. Reid

*Subroutine YWPR has a problem unless your computer automatically sets variables to zero before they are used. Its statement 3 says DP=DP+G(NN+1−K)*DPHI(K+1), but DPHI(K+1) is not defined when K=NN. DP is never used when K=NN, so the simplest solution is to set DPHI (LG+1) = 0 just after statement 1 which defines the lower index values of DPHI.

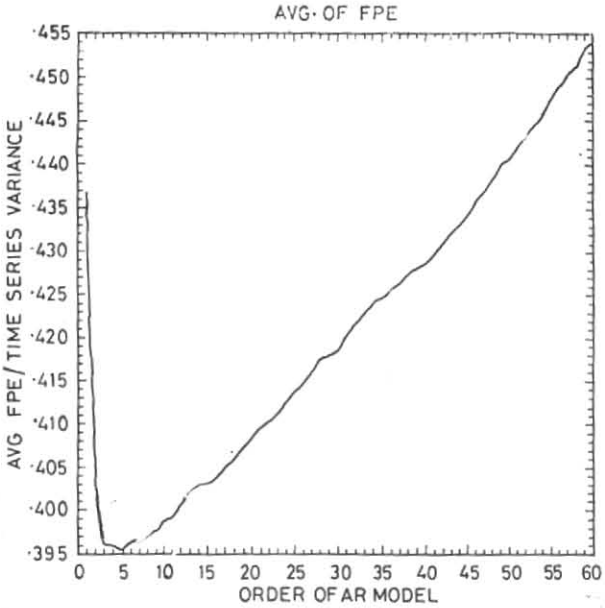

Fig. 1. Average of Akaike's Final Prediction Error normalized by dividing by the time series variance, plotted as a function of the AR model fitted to time series. The data on which this plot is based are 27 summers (1 June-28 October 1951-1977) of zonal wind data at ·85 kPa over New Delhi

(1979) has proposed a method to estimate confidence limits for MEM spectra, but we do not understand his procedure well enough to apply it. Instead, we will discuss statistically sound methods for confidence limits that are easy to understand and apply.

We consider two cases. The first case is where an observed time series is long enough that it can be cut into 10 or more pieces which can be analyzed independently. This is the situation with our radiosonde data where we have between 10 and 25 summers of data for each of the cities under study. The second case is where a time series is too short to be split into pieces.

When a time series is long enough to be split into ten or more separate pieces, the simplest way to estimate confidence limits for MEM spectra is analogous to what Bartlett proposed for traditional spectral analysis (*see* the examples in Jenkins and Watts 1968, p. 239). Cut the observed time series into 'M' pieces of equal length and find the average estimated spectrum :

$$\bar{S}(f) = \frac{1}{M} \sum_{m=1}^{M} S\,(f; m); \qquad (3)$$

where $S\,(f;\, m)$ is the estimated MEM spectrum of the m^{th} piece of the time series. By the Central Limit Theorem, if 'M' is sufficiently large (greater than 10 is often sufficient), then the variance of $\bar{S}\,(f)$ at frequency f will be nearly equal to σ_f^2/M where σ_f^2 is the variance of $S\,(f\,;\, m)$ across the different values of m at fixed frequency f. The centred 90% confidence limits would be $\bar{S}\,(f) \pm 1.65\sigma_f/\sqrt{M}$, where both $\bar{S}(f)$ and σ_f are functions of frequency.

When a time series is too short to be cut into 10 or more pieces, we can still estimate the variability of an estimated spectrum. Suppose that we have estimated an AR model of a specified order N based on a time series of length T, and we want to know how different the spectrum would look if we had observed different realizations of that random process. To answer this question, we use the AR coefficients determined from the observed time series along with a Gaussian random number generator to generate 'M' simulated time series of the form of Eqn. (1), where 'M' is several hundred or a thousand. In order to compute each simulated time series, we must arbitrarily choose N start-up values $\{x_1, \ldots, x_N\}$. Each of the M simulated time series should be sufficiently longer than T so that they can be truncated to length T by dropping enough values from the beginning of the time series so that all "memory" of the precise start-up values will be lost. This means that we need to omit $N + L$ points from the beginning of each simulated time series, where N is the order of the model and where L is the number of lags needed for the autocorrelation of the model to go essentially to zero.

For each simulated time series truncated to length T, we estimate an AR model. Using (2), we compute the model's spectrum $S(f_j\,;\, m)$ at a set of discrete frequencies f_j, for $=1, \ldots, J$; $m=1, \ldots, M$ is the

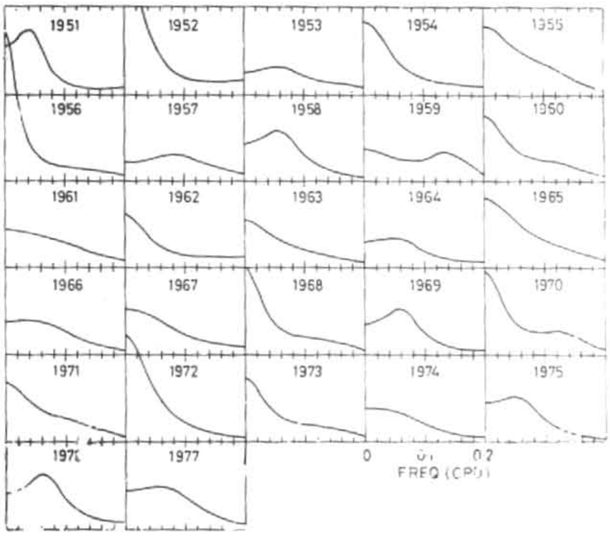

Fig. 2. MEM spectra based on 150-day time series (1 June-28 October) for 85 kPa zonal wind over New Delhi. Years covered: 1951-1977. Tenth order AR models were fitted to the summer time series. The value at the top of each spectral density axis is 400 $(m/s)^2/(cycles/day)$.

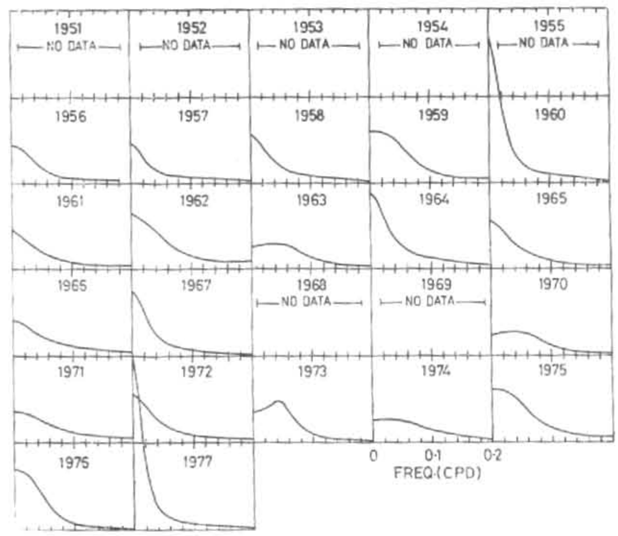

Fig. 3. MEM spectra as in Fig. 2 but for 85 kPa zonal wind over Trivandrum

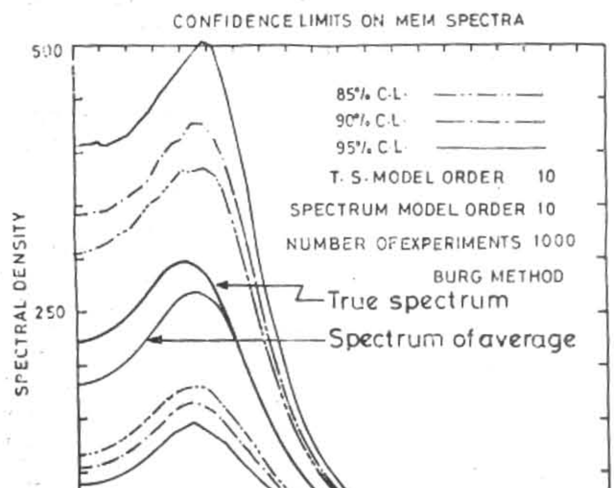

Fig. 4. 1000 time series were generated from the tenth order AR model that produced the spectrum for 1951 in Fig. 2. The spectrum of this model is referred to above as the true spectrum. The spectrum of the average of the 1000 tenth order AR models fitted to the generated time series is also shown above with its confidence limits, which indicate the spread in the values of the 1000 spectra. Frequency units are (cycles/day). Spectral density unit are $(m/s)^2/(cycles/day)$.

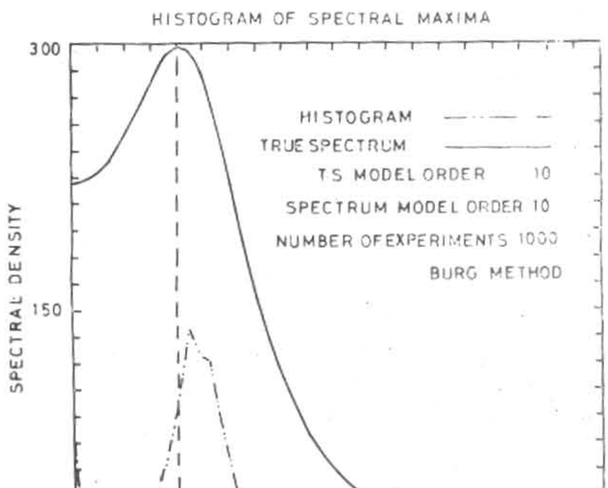

Fig. 5. The 1000 spectra summarized in Fig. 4 were computed at $f_j = (j)$ (0.004 cylces/day) for $j = 0, \ldots, 50$, Let n_j be the number of spectra that had their maxima at f_j. The histogram depicting n_j is plotted as a dashed curve. For reference, the spectrum of the process used to generate the 1000 time series is plotted as a solid line. The numbers labelling the ordinate apply both to n_j as well as to the spectral density, whose units are $(m/s)^2/(cycles/day)$.

counter for the simulated time series. Centred 90% confidence limits at each fequency f_j are then found by fixing f_j and sorting through the m-index of $S(f_j ; m)$ to find the values in the 5th and 95th percentiles.

3. MEM spectra summer by summer

We begin by plotting in Fig. 1 the average of Akaike's Final Precition Error (FPE) as a function of AR model order. The average is based on AR models fitted to 27 summers (1951-1977) of 85 kPa zonal wind data for New Delhi. Ulrych and Bishop (1975) suggest that the order of the AR model be chosen so that the FPE is minimum. The average FPE is minimum for a fifth order model, and the authors have found similar results when fitting AR models to monsoon data from NMC operational analyses. Nonetheless, we decided to use tenth order AR models for our analyses to minimize possible loss of spectral structure. As the reader will see, even with 10th order AR models, these particular spectra do not display much fine structure. Picking a model order is a compromise between competing factors. If too high an order is chosen, MEM can produce spurious spectral peaks (Ulrych and Bishop 1975, pp. 188-189; Reid 1979, p. 5289). Conversely, precise representation of a strictly periodic time series with period 50 days would require a 100th order AR model if the interval between observations is one half day. This is the standard problem with spectral analysis : one cannot have both high resolution in frequency as well as stable estimates of the spectral density function.

Spectra based on 10th order AR models for the 85 kPa zonal wind component at New Delhi and Trivandrum are shown in Figs. 2 and 3, respectively. The results for other cities are similar. Although time scales longer than 10 days explain most of the variance, most years do not have identifiable spectral peaks at 10-20 or 30-50 day time scales. Rather, red noise predominates. Fig. 4 shows the results from simulating $M=1000$ summers based on the 10th order AR model for 1951 which did display a low frequency peak. The average estimated spectrum from the simulations closely recovers the spectrum of the process used to generate the simulations, although the confidence limits are fairly broad. Perhaps more important is Fig. 5, which tabulates the frequencies at which the 1000 spectra computed from the simulated time series had their spectral maxima. Roughly 95% peaked at a frequency near the peak in the 1951 spectrum, and less than 5% peaked at zero frequency, which would be the case with red noise.

From this, we conclude that the observations are more consistent with red noise than with a 10th order AR model that has a low frequency peak. In particular, these spectra do not separate intraseasonal variability into 10-20 and 30-50 day time scales. The same result applied to all twelve cities we analyzed. It is possible, though, that Akaike's FPE is not the best way to choose the model order and/or that principal components of EOFs computed from spatially distributed data might separate 10-20 and 30-50 day time scales. To examine these possibilities, we plan to further investigate the properties of MEM spectra and to compute EOFs and their principal components.

4. Time-height diagrams of zonal wind

We filtered time series with a 30-50 day filter in order to examine the vertical structure of activity on this time scale. Over all cities, fluctuations occur throughout a deep layer in the troposphere. At New Delhi, this extends all the way to 10 kPa; at Trivandrum to 30kPa. Over both these cities, there is little vertical propagation. At 30-50 day scales in the NMC operational analyses for 1979, Mehta and Ahlquist (1986, p. 169) reported little vertical propagation at 20°N but there was upward propagation at the equator. The record for Calcutta exhibits both upward and downward propagation.

5. Summary

We regard the following as the main findings of our research to date :

(*a*) Confidence limits for MEM spectra can be estimated with statistical methods.

(*b*) Intraseasonal time scales are important in the summer monsoon throughout the troposphere, but MEM spectra based on 10th order AR models at individual cities do not separate 10-20 and 30-50 day timescales. This finding may reflect reality, or it may result from an overly conservative choice of model order (even though Akaike's FPE recommends a lower order than what was used), or spatial scales may need to be separated in order to cleanly separate 10-20 and 30-50 day activity.

(*c*) Vertical structure of 30-50 day activity over India during the summer is in phase throughout a deep tropospheric layer. Vertical propagation occurs over Calcutta but is not prominent over New Delhi and Trivandrum.

As follow-up work to this study, we plan to investigate horizontal structure on intraseasonal time scales. This will include principal component analysis and time sequences of horizontal and vertical structure.

Acknowledgements

This research was sponsored by NSF grant ATM-8714674. The radiosonde data were obtained from the National Center for Atmospheric Research, which is sponsored by the National Science Foundation. Simulations to deterimine confidence limits for MEM spectra were performed using Cyber 205 computer time made available by the Supercomputer Computations Research Institute at Florida State University.

References

Bendat, J.S. and Piersol, A.G., 1986, *Random Data Analysis and Measurement Procedures,* Second edition, Wiley and Sons, New York, p. 404.

Box, G.E.P. and Jenkins, G.M., 1976, *Time Series Analysis : Forecasting and Control,* Holden-Day, San Francisco.

Jaynes, E.T., 1982, On the rationale of maximum-entropy methods *Proc. IEEE,* **70**, 939-952.

Jenkings, G.M. and Watts, D.G., 1968, *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 525 pp.

Mehta, A.V. and Krishnamurti, T.N., 1988, Interannual variability of the 30 to 50 day wave motions, *J. met. Soc. Japan*, **66**, 535-548.

Mehta, V.M. and Ahlquist, J.E., 1986, Interannual variability of the 30-50 day activity in the Indian summer monsoon, *Met. atmos., Phys.*, **35**, 166-176.

Papoulis, A., 1981, Maximum entropy and spectral estimation : A review, *IEEE Trans. on Acoustics, Speech & Signal Processing*, *ASSP*-**29**, 1176-1186.

Reid, J.S., 1979, Confidence limits and maximum entropy spectra *J. geophys. Res.*, **84**, 5289-5301.

Ulrych, T.J. and Bishop, T.N., 1975, Maximum entropy spectral analysis and autoregressive decomposition, *Reviews of Geophysics & Space Physics*, **13**, 183-200.

van den Bos, 1971, Alternative interpretation of maximum entropy spectral analysis, *IEEE trans. Info. Theory*, *IT*-**17**, 493-494.