1. Publication Nº	2. Version	3. Date	5. Distribution
INPE-2253-PRE/036	*Second*	*Oct., 1981*	☐ Internal ☒ External ☐ Restricted
4. Origin *DME*	Program *SINO*		

6. Key words - selected by the author(s)
MODELING BY MARKOV PROCESSES *SPECTRUM ANALYSIS*
PERSISTENCE IN DAILY DATA
NONSTATIONARITY

7. U.D.C.: *551.511.61*

8. Title *INPE-2253-PRE/036*

A STUDY OF METEOROLOGICAL TIME SERIES.
PART I: MODELING PERSISTENCE

10. Nº of pages: *42*

11. Last page: *38*

12. Revised by

Ivan Kantor

9. Authorship *Srinivasan Srivatsangam*
Vernon Edgar Kousky

Responsible author

13. Authorized by

Nelson de Jesus Parada
Diretor

14. Abstract/Notes

The role of persistence in time series can be modeled by autoregressive (or Markov) processes, provided that the series are stationary. In general, meteorological variables have a pronounced seasonal cycle, which makes the mean a function of time. Hence, the seasonal cycle must be removed before stochastic modeling, or spectrum analysis, is done. Retention of the seasonal cycle leads to the false conclusion that a first order autoregressive model best represents meteorological time series. If the seasonal cycle is removed before modeling, the best fits are usually given by higher order autoregressive processes. These higher order processes quite often have free solutions which are oscillatory and, therefore, are more relevant to the dynamics of the weather.

15. Remarks

To be submitted for publication in Monthly Weather Review.
Revised August 1982.

2253

A STUDY OF METEOROLOGICAL TIME SERIES
PART I: MODELING PERSISTENCE

S. Srivatsangam
and
Vernon E. Kousky

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO - (CNPq)
INSTITUTO DE PESQUISAS ESPACIAIS - (INPE)
12200 - SÃO JOSÉ DOS CAMPOS, SP, BRAZIL

Revised August

RESUMO

O comportamento de persistência em séries de tempo pode ser modelado através de processos autoregressivos (ou Markovianos), contanto que as séries sejam estacionárias. Em geral, variáveis meteorológicas têm um ciclo sazonal pronunciado, que resulta na média depender do tempo. Assim, o ciclo sazonal deve ser filtrado, antes de se proceder à modelagem estocástica ou à análise espectral. Manter o ciclo sazonal conduz à conclusão falsa que o modelo autoregressivo de primeira ordem é o que melhor representa as séries temporais em Meteorologia. Se o ciclo sazonal é removido antes da modelagem, os processos autoregressivos mais apropriados são os de ordem maior que um. Frequentemente, estes processos têm soluções livres que são oscilatórias e, portanto, são mais relevantes à dinâmica do tempo.

Abstract

Autocorrelation functions (acf's) and spectra are computed for many meteorological time series consisting of direct observational data and modified data. Time series of unmodified data have acf's which often decay very slowly with lag (counted in days), and are modeled correctly by an autoregressive (AR) model of order 1. This low-order model indicates great persistence and produces a red-noise spectrum.

Meteorological time series contain information about phenomena such as cyclones, anticyclones and blocking, as well as about the annual cycle and other slow oscillations such as the semiannual cycle and climatological variations. Since the latter have longer periods than the former, it is of interest to study time series not containing the latter, with a view to assessing the statistical characteristics of the relatively high-frequency weather phenomena. When meteorological time series modified for this purpose are studied, the acf's are seen to decay rapidly. The appropriate models for the modified time series are AR processes of order 2 or more. These higher order models often have dynamical representation in terms of oscillating mechanisms, which are more characteristic of weather systems.

1. Introduction

Meteorological events occur in space and time. Hence, to comprehend them,it is necessary to analyze meteorological data in both these domains.

Spatial analysis of meteorological data is made somewhat easier than temporal analysis by the fact that the former is in a finite domain. Time is unbounded and time series analysis involves fundamental assumptions concerning determinism, stationarity, etc., which demand greater initial care than spatial analysis.

The question of the time scale of meteorological events is, however, unavoidable. This scale must be distinguished from the climatological time scale and also from the scale of the seasons.

An aim of this work is to attempt a separation of these scales and, thus, isolate a frequency range which may be deemed specifically meteorological. This separation of scales is aided by the conventional power spectrum analysis as well as by stochastic modeling, which is not as widely applied.

The stochastic model, various types of which are discussed at length by Box and Jenkins (1970), becomes useful in representing meteorological series, since it could be made to incorporate the memory or persistence inherent in the series. The stochastic process which exhibits this quality is the autoregressive process of order p, denoted by AR(p). If x is the variable whose changes are to be treated, we write this process as

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots\cdots + \alpha_p x_{t-p} + a_t \tag{1}$$

Here the relative observational time is subscripted and a_t is a random (or white noise) process, whose theoretical mean, variance and autocorrelation function (acf) are given, respectively, by

$$E\,[\,a_t\,] = 0 \tag{2}$$

$$\mathrm{var}\,[\,a_t\,] = \sigma_a^2 \tag{3}$$

and

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{4}$$

In the last of the above equations, k denotes the k^{th} lag. In writing (1) it has been assumed that the time average of x has been removed.

The expression (1) is a p^{th} order regression of x upon itself. If all the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_p$ are zero, the process x_t is equivalent to the white noise a_t, which has no memory as seen from (4). If, however, the coefficients in (1) are non-zero, x_t is said to have memory. Whereas the white noise a_t has a uniform distribution of spectral energy with frequency, the processes with memory have nonuniform distribution of energy; and the nonuniformity of distribution is a function of the acf. This is because the coefficients α_k of (1) can be determined from the autocorrelation coefficients ρ_k.

As the spectrum is just the Fourier transform of the autocorrelation function, it is transparent that the autoregressive model and the spectrum are closely connected (see, e.g., Childers, 1978). It is also obvious that the structure of the acf would affect both the selection of a scientifically informative or useful AR model and the spectral density distribution.

In view of these considerations, it is necessary to evaluate the time series *a priori*, for instance using physical considerations, as to their various parts such as trend, seasonality, and fluctuating (or stochastic) components (see Parzen, 1974). This is so because both trends and seasonal components can inflate the acf positively such that the acf does not die out rapidly (see Box and Jenkins, 1970, p.175). Now, what is called trend in statistical parlance may be short lived in a long time series, and therefore it is legitimate to think of long period oscillations as nonlinear trends.

One may deem the seasonal cycle as being deterministic (see Jones, 1964) and other low-frequency oscillations in meteorological time series as being stochastic. The seasonal cycle, therefore, in addition to inducing a nondamping quality to the acf due to the high-frequency components alone, produces nonstationarity. This nonstationarity means that the mean or the statistical expectation is dependent on the calendar day. As the conventional procedures for time series analysis, both modeling and spectrum analysis, are valid only for stationary time series by assumption, it is doubly necessary to remove the seasonal cycle.

If the seasonal cycle is not removed from meteorological time series, the acf is often so positively augmented that an autoregressive model of the first order, AR(1), is chosen. This model has

$$\rho_k = \rho_1^k \tag{5}$$

With $\rho_1 > 0$, the acf of AR(1) can never be less than zero. The spectrum of AR(1) reflects this by having generally higher spectral density at lower frequencies (see Gilman *et al*, 1963). Such a distribution of spectral energy is known as "red noise" (see Lorenz, 1965). It may be noted that whereas AR(1) always has a red-noise spectrum, higher order AR processes may or may nor have red-noise spectra. AR(p) of order greater than 1 may have more rapid damping of the acf than is indicated by (5), resulting in stochastic spectral plateaus at some frequency intermediate to 0 and the Nyquist frequency (see, e.g., Box and Jenkins, p. 63).

Until rather recently, a widely accepted stochastic model for meteorological time series has been AR(1) (see Lorenz, 1965, Julian, 1971 and Leith ,1975). The study of Straus and Halem (1981), however, concludes that at least a second-order AR process is required to adequately describe surface observations and simulated data from a general circulation model.

Our aims in this paper are to demonstrate how, by biasing the series through the deletion of the seasonal cycle and other low-frequency components, one might arrive at AR(p) models with $p > 1$, and to show the fundamental dynamical differences between AR(1) and certain kinds of higher order AR processes.

2. Data, method of analysis and terminology

We intend to determine here the common properties of meteorological time series. With this in mind, we have brought together a wide variety of variables. The variables are a zonal index, a meridional index, station pressure and 850-mb temperature. They pertain to the tropics and extratropics of both the Northern and Southern Hemispheres.

The upper air data used have been taken from the U.S. National Meteorological Center tropical grid analyses, which were obtained from the National Center for Atmospheric Research* at Boulder, Colorado. Only 1200 GMT data are used in this study. The missing data, which amounted to only nine days in the four-year period from 1 July 1975 to 30 June 1979, were linearly interpolated.

The Brazilian station pressure data were obtained from the Instituto Nacional de Meteorologia in Brazil. Observations for 1200, 1800 and 0000 GMT were averaged to give a single value for a day. These values are used to form the time series of station pressure studied here.

We have stationarized the mean of each time series by removing Fourier components, in the manner of Blackmon (1976) and Blackmon *et al.* (1977). For a series one year in length, the first four harmonics are subtracted out. The sum of these four harmonics is referred to as the estimated seasonal cycle.

The autocorrelation function estimates are positive definite rather than unbiased. That is, if the series has N observations, the k^{th} lag correlation is not computed by using N-k as the divisor, but, rather, N itself. This produces estimates with smaller mean square error, according to Jenkins and Watts (1968, p. 184). Jones and Kearns (1976) have mentioned that positive definite estimates de-emphasize the higher lags.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

A lag window is always used in obtaining the spectrum. The rectangular window may be chosen intentionally or by default. It may be seen from Jenkins and Watts (1968, p. 252) that, for the same time series and length of acf used, the rectangular window has fewer degrees of freedom than the Bartlett, Tukey and Parzen windows. As a result, the rectangular window produces less biased estimates, which, however, have greater variance. Since we are not intent on discovering any spectral peaks, we decided to use the Parzen lag window. This is a wide window which yields biased, but less variable, estimates.

The smoothed spectral density estimates, g(f), were obtained through the discrete cosine transform of the window-modified autocorrelation function:

$$g(f) = 2\Delta t \left\{ 1 + 2 \sum_{k=1}^{L-1} r_k \, w_k \cos(2\pi f k \Delta t) \right\} \tag{6}$$

In (6) the frequency, f, has the range $0 \leq f \leq 1/(2\Delta t)$, Δt being the time span of unit lag. Also, r is the estimated autocorrelation function, w the Parzen window function and L the maximum lag.

The maximum lag, L, we have used is 40. Since the shortest of our time series are daily data for a year, having N = 365 (366, in the case of leap year), L is never more than a ninth of N. This is in accord with a Jenkins and Watts caveat to utilize fewer lags, as a large L would reduce the degrees of freedom, narrow the bandwidth and, possibly, generate spurious peaks.

Following a suggestion of Jenkins and Watts (1968, p. 255), we have plotted the spectra on a logarithmic scale. This has the advantages that more detail is seen over a wider ordinate range and that the confidence limits become independent of frequency.

The degrees of freedom for the Parzen window are given by

$$\nu = 3.71\ N/L. \tag{7}$$

With this equation and the information contained in the Jenkins and Watts text, we have indicated the confidence intervals and bandwidths for the estimated spectral densities.

The older methods of computing the spectra and related statistics are used instead of the more recent techniques collected together in Childers (1978). We believe that more meteorologists would be familiar with the older methods. Hence, their use will facilitate understanding of our findings. Moreover, since our time series are long, containing more than a hundred observations, the use of the maximum entropy method, or other methods, meant especially for shorter records, is not crucial.

In modeling, rather than using the final prediction error criterion of Akaike (see Childers, 1978), we have used the straightforward variance reduction method. For long time series, the two methods must yield nearly the same results, as can be verified from Ulrych and Bishop (1975).

The coefficients of (1) are computed from the autocorrelation function via the Yule-Walker equations (see, e.g., Box and Jenkins, 1970; Parzen, 1974).

3. Results

The first of our illustrations, Fig. 1, pertains to the zonal index at 41°N and 500 mb for the year 1978. We define the zonal index, I_z, as

$$I_z = (1/72) \sum_{i=1}^{72} u_i \tag{8}$$

where the zonal component of the wind at a particular longitude is denoted by u_i, data being available at 72 longitudes, at 5 degree intervals.

In Fig. 1, the autocorrelation function (acf) of the unmodified zonal index has a unit lag value greater than 0.9. The acf decays extremely slowly with lag, being 0.255 even at 40 lags. It is, of course, true that the autoregressive process of order one, i.e., AR(1), has a monotonically decreasing acf, if the unit-lag correlation is positive. This can be verified from (5). However, the decay of the acf in Fig. 1 is not as rapid as that calculated from (5). Moreover, we know that a nondecaying acf is indicative of nonstationarity (Box and Jenkins, 1970, p. 55). The prime suspect then is the seasonal cycle, on the back of which, so to say, the weather is carried.

In Table 1, the percent variance explained by the harmonics contained in the estimated seasonal cycle is given. In the case of the zonal index, the annual cycle alone explains over 60% of the variance, and the remaining three harmonics together explain another 4.5%. There can be little question that the annual cycle is deterministic. It would indeed be wrong to treat this cycle as part of the statistical variations.

It is arguable that the second, third and fourth harmonics of the annual cycle are not statistical, but, rather, deterministic. Nevertheless, following the precedent set by Blackmon (1976), Blackmon *et al.* (1977) and the subsequent publications of these authors, we have chosen to discard these higher harmonics of the annual cycle as well.

When the estimated seasonal cycle is subtracted out, the residuals yield an acf, shown in Fig. 2, that decays much faster than the acf of the raw zonal index data. For this series of residuals, the first negative value of the acf is at 12 lags. Since the unit-lag correlation here is about 0.8, it can be calculated from (5) that the corresponding AR(1) must have $\rho_4 \cong 0.4$, $\rho_8 \cong 0.16$ and $\rho_{10} \cong 0.1$, values which are in reasonable agreement with the estimated values in Fig. 2. This agreement is again seen between the smoothed spectral density estimates and the theoretical AR(1) spectrum given in Fig. 3.

The residual variances (i.e., the square of the standard errors) for AR models up to order 3 are given in Table 2. It may be recalled that the residual variance is not a monotonically decreasing function of the order (or "predictors" included; see Haan, 1977, p. 210). The residual variances are the same for the AR(1) and AR(2) models of the zonal index. Since AR(1) is the more parsimonious model, having fewer parameters, one might accept it as the best model for the series of zonal index with stationarized mean.

The time variations of the zonal index may be thought of as second-order effects of atmospheric eddies (see, e.g., Phillips, 1954, 1956). It is, therefore, worth analyzing some variable representing the eddies directly. The meridional index I_m is such a variable. We define it as an area average of the square of the meridional component of the wind at 500 mb. The boundaries of the area we have chosen are the latitudes 28.7°N and 48.1°N and the meridians 60°W and 120°W. Then it can be written

$$I_m = (1/78) \sum_{i=1}^{13} \sum_{j=1}^{6} v_{ij}^2. \tag{9}$$

In (9), i is the longitude index and j, the latitude index.

For the meridional index and the remaining variables in this Part, we show the estimated spectral density distributions only after the removal of the estimated seasonal cycle. The spectral density estimates for the meridional index are illustrated in Fig. 4. The estimates for the station pressure at Manaus and São Paulo in Brazil are in Fig. 5. these estimates are for the original series less the first two harmonics of the annual cycle. Finally, estimates for 850 mb temperature at three longitudes intersecting 41°N are in Fig. 6. The AR(2) models, and in some instances the AR(1) model as well, are shown in Figs. 4, 5, and 6. The AR(2) model is seen to represent well the overall features of the estimates.

This is further confirmed by the data on residual variances in Table 2. These data indicate that for all the variables, except the zonal index, the best fitting AR models have order higher than 1, provided that the series are first stationarized for the mean by removing harmonics of the annual cycle. It is true that the higher order models often explain only a small fraction of the residual variance of AR(1). From the point of view of parsimony alone, it may be thought preferable to select the AR(1) model. It will be shown in the next section that parsimony here is a dubious criterion.

In Table 3 we present the acf's for both the original series and series modified by the subtraction of harmonics of the annual cycle. The former possess acf's which decay slowly, indicating nonstationarity of mean as suggested by Parzen (1974) and Box and Jenkins (1970, p. 175). In the latter series this sluggishness is gone. That is, when the estimated seasonal cycle is removed, each modified series has a rapidly decaying acf structure at the first few lags; thereafter the acf fluctuates about zero. Thus, except at the first few lags, the acf's of the modified series are almost like the acf of white noise, as given by (4). The modified series quite obviously have much less persistence than the original series.

It may be argued that if the seasonal cycle is estimated with multi-annual data, the estimates will be weaker than those made with a single year's data, thus affecting our conclusions. To explore this, we present the acf's for 850 mb temperature at various geographical positions, using daily data for the 4 year period from 1 July 1975 to 30 June 1979. Acf's are shown for the original series as well as those modified in different manners. It is seen from Table 4, which contains these acf's, that the annual cycle inflates the acf more than any other low-frequency component in almost all cases. Moreover, the removal of the annual cycle decreases r_2 much more than r_1, thus enforcing a rapid decay of the acf. The subtraction of the other low-frequency components — which may be thought of as a filtering procedure — merely augments the tendencies set by the removal of the annual cycle. That is, the general filtering of low-frequency components merely ensures that the models of the filtered series are higher order AR processes with a special characteristic, as will be seen from the ensuing discussion.

4. Discussion

We need to justify the choice of modeling persistence effects alone, instead of commencing, say, with autoregressive-moving average models, as elucidated in the Box and Jenkins book. The choice is justified partly by the historical precedents, mentioned above, which recognize the persistence in meteorological time series. Partly it reflects our agreement with Kashyap and Rao (1976): "Accurate estimation of parameters in a system involving moving average terms is considerably more difficult than the estimation problem in a system without moving average terms .... Consequently, in modeling empirical time series, the possibility of developing a model free of moving average terms should be explored so that the model satisfactorily represents the data".

The initial omission of the moving average process signifies that we do not have an *a priori* reason to believe that the random processes affecting the weather are in turn persistent. This is seen from the equation for the q^{th} order moving average process

$$a_t = \Theta_1 a_{t-1} + \Theta_2 a_{t-2} + \dots + \Theta_q a_{t-q} - x_t \qquad (10)$$

where x_t is the process under study. Comparison may be made between (10) and (1).

However, the best reason for choosing a model for the weather from the class of autoregressive models lies in the following physical consideration.

As can be deduced from (1), autoregressive modeling always leads to a finite difference equation. This equation is inhomogeneous, and it is vital to consider the free solutions corresponding to the homogeneous part, as well as the forced solution. The free solutions lead to insights concerning the mechanism which is given as a mathematical expression, while the forced solution tells us how this mechanism functions under the specified forcing. Moreover, the characteristics of the free solutions are reflected in the acf, which is central to much work on time series analysis.

The homogeneous difference equation corresponding to AR(1) is

$$x_t = \alpha_1 x_{t-1} \tag{11}$$

where $\alpha_1 = r_1$, the unit-lag correlation coefficient. The solution of (11) is

$$x_t = K\alpha_1^t \tag{12}$$

where K is the value of x_t at $t = 0$. It must be remembered that t is always a positive integer or is zero.

From the information on acf given above we have $0 < \alpha_1 = r_1 < 1$. So, for all these series, the process (10) is convergent. Moreover, as $\alpha_1 > 0$, the free solution of our AR(1) process is a convergent monotonic process (see Gandolfo, 1971).

Then it may be asked if the free response of any part of the atmosphere is ever dynamically treated as a monotonically increasing or decreasing function. Certainly the reply to this question is in the negative.

Apparently, the homogeneous part of AR(1) can represent trend or long-period oscillations, and not the weather.

The homogeneous part of AR(2) is

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} \tag{13}$$

for which let us assume a solution of the form

$$x_t = m\, s^t \tag{14}$$

which resembles(12). Substituting (14) into (13), we will have a solution if s is a root of

$$s^2 - \beta_1 s - \beta_2 = 0 \tag{15}$$

The two roots of (15) may be real and distinct, provided that

$$\beta_1^2 + 4\,\beta_2 > 0 \tag{16}$$

Then, the two independent general solutions are

$$x_t^{(1)} = m_1\, s_1^t \quad \text{and} \quad x_t^{(2)} = m_2\, s_2^t$$

which may be combined to give

$$x_t = m_1 s_1^t + m_2 s_2^t \tag{17}$$

The coefficients m_1 and m_2 are determined by the initial conditions.

The convergence of (17) is assured if $|s_1|$ and $|s_2|$ are both less than unity.

If both s_1 and s_2 are greater than zero, (17) produces the same result as (12), representing trend. If, however, say, $s_1 > 0$ but $s_2 < 0$, rather interesting departures from (12) appear. The solution corresponding to s_2 must then change sign with each increment in t, a phenomenon that Gandolfo (1971) calls improper oscillation. This improper oscillation would modify the monotonic growth or decline arising from the solution with s_1.

If, in (15)

$$\beta_1^2 + 4\,\beta_2 < 0 \tag{18}$$

the roots s_1 and s_2 are complex conjugates. The solution may be written as

$$x_t = m_1 S^t \; e^{i2\pi f_0 t} + m_2 S^t \; e^{-i2\pi f_0 t} \tag{19}$$

where

$$S = (-\beta_2)^{\frac{1}{2}} \tag{20}$$

and

$$f_0 = (1/2\pi)\arccos \left\{\beta_1/(-4\,\beta_2)^{\frac{1}{2}}\right\} \tag{21}$$

The oscillatory solution (19) is discontinuous in time as t is discrete. The convergence or divergence of (19) depends on whether $|\beta_2| < 1$, or $|\beta_2| > 1$. If $\beta_2 = -1$, the solution is purely oscillatory with neither amplification nor damping.

The coefficients β_1 and β_2 are computed via the Yule-Walker equations (see Box and Jenkins, 1970) as

$$\left.\begin{aligned} \beta_1 &= r_1 \; (1 - r_2) \; / \; (1 - r_1^2) \\ \beta_2 &= (r_2 - r_1^2) \; / \; (1 - r_1^2) \end{aligned}\right\} \tag{22}$$

where r_1 and r_2 are estimates of the unit lag and second lag autocorrelation coefficients, respectively. Table 5 has the values of β_1 and β_2 for the series under discussion, indicating that, for all of them, $\beta_1 > 0$ and $\beta_2 < 0$.

From Table 5 it is apparent that all of the series analyzed, excepting the zonal index, satisfy (18), once the seasonal cycle is removed. So these series have AR(2) models with free oscillatory solutions. These solutions are also damped with time, as $|\beta_1| < 1$ for the series, as exhibited in Table 5.

Since the majority of the data treated have AR(2) models whose homogeneous parts have oscillatory solutions, it may be asked if the free response of the atmosphere is thought of as being oscillatory. The answer has been affirmative ever since the publication of Rossby (1939). Although the existing literature confirming the undulatory property of the weather is scanty, energy propagation by Rossby waves, as given by Hovmöller (1949) and Parry and Roe (1952) may be adduced as partial evidence.

The solution for the inhomogeneous equation for AR(1), with $\alpha_1 > 0$, is marked by a concentration of spectral energy in the low frequency portion of the spectrum. This feature is dealt with at length by Gilman *et al.* (1963), Jenkins and Watts (1968) and Box and

Jenkins (1970). There is no need to consider it again here.

The formal solutions for all orders of autoregressive processes are given by Box and Jenkins (1968, p. 54), and Jenkins and Watts (1968). We will present here only a few points not explicitly dealt with by them.

Box and Jenkins have shown that the acf's of the autoregressive processes satisfy difference equations analogous to the homogeneous equations examined above. Therefore, if the free solutions of, e.g. AR(2), are damped and oscillatory, then so is the acf. The damping factor and the frequency of the acf are, accordingly, identical to (20) and (21) for the free solutions.

The spectrum, however, being the cosine transform of the acf, may be expected to display some special characteristics of its own, when the acf has a damped oscillatory structure. This proves to be true.

The spectrum of AR(2) is given by

$$\left.\begin{aligned} p(f) &= 2\,\sigma_a^2 \Big/ \left\{1 + \beta_1^2 + \beta_2^2 - 2\beta_1\,(1 - \beta_2)\cos 2\pi f - 2\beta_2 \cos 4\pi f\right\}, \\ &\qquad\qquad 0 \le f \le 1/2 \end{aligned}\right\} \quad (23)$$

Upon differentiating (23) with respect to frequency f, and setting the resulting expression equal to zero, we get an equation for the frequency having an extremum in the spectrum

$$f_1 = (1/2\pi) \arccos \left\{ \beta_1 (\beta_2 - 1)/4\ \beta_2 \right\} \tag{24}$$

The frequency f_1 is different from the frequency of the free oscillation and of the acf, given by f_o in (21). We can write (24) as

$$f_1 = (1/2\pi) \arccos \left\{ \left[\beta_1/(-4\ \beta_2)^{\frac{1}{2}}\right] \left[(1 - \beta_2)/(-4\ \beta_2)^{\frac{1}{2}}\right] \right\} \tag{25}$$

The expression within the first pair of brackets of (25) alone will lead to f_o, if

$$(1 - \beta_2)\ /\ (-4\ \beta_2)^{\frac{1}{2}} \tag{26}$$

equals unity, that is, if $\beta_2 = -1$. As β_2 rises from this value and approaches zero, the denominator of (26) becomes increasingly more dominant, and, as can be seen from (25), f_1 decreases with respect to f_o. One also infers from (20) that as $\beta_2 \to 0$, the amplitude of the oscillation weakens. In fact, the expression in (26) may be so large that the entity within the braces of (25) overshoots the functional bound at 1; following this, there will be no f_1 although f_o has a definite value. This is plainly a reversion to AR(1). Reference may be made to Papoulis (1962) and Godin (1972) for further clarification of the issues arising here.

In Table 5, the values of β_2 are larger than -1, but less than zero. Obviously, where (18) is satisfied, the spectral peak occurs at a lower frequency than the free oscillation. That is, $f_1 < f_o$. And, of course, twin-sided spectral peaks do not always occur when the free solution is oscillatory, as can be seen from Table 5. This can be further verified from the theoretical AR(2) spectra for station pressure at Manaus and 850 mb temperature at 41ºN, 120ºW. This contrast of the acf and the spectrum makes for their essential difference.

The plateaus in the spectra of AR(2) reflect the tendency of the free solutions to be oscillatory. Hence the term 'pseudo-periodic behavior', employed by Box and Jenkins (1970) in this context, seems to be misleading. Perhaps 'AR(2) with complex roots' is adequate, as it reflects the cause of oscillatory response found in other parts of dynamics, for instance, mechanical vibrations (Tse *et al.*, 1963).

We have discussed the AR(2) process at some length, not because it is always the best model for meteorological time series, but rather because the chief difference between AR(1) and all higher order autoregressive models is one and the same thing. It is that AR(1) can never have free oscillatory solutions while all higher order processes permit such solutions.

5. Conclusion

In this paper we have attempted to assess the influence that relatively low-frequency components, such as the seasonal cycle and climatological oscillations, have on meteorological time series, by studying the time series with and without such low-frequency components. Filtered time series containing only the high-frequency oscillations have, at lags of one and two days, positive autocorrelation. This indicates the persistence in the series. But these lag correlation coefficients are both small and rapidly decaying with lag. This is characteristic of stochastic processes which have a low degree of persistence.

However,unfiltered time series containing low-frequency oscillations have lag correlation coefficients which are large and positive up to lags of several days. Therefore, unfiltered time series exhibit very strong persistence. A chief disadvantage of using unfiltered time series is the possibility of mistakenly attributing their strong persistence to the weather itself. Furthermore, it is essential to remove the seasonality in the time series, for otherwise there would be nonstationarity in the mean (or statistical espectation) and the conventional methods of time series analysis are concerned with stationary series only (Jones,1971).

The stationarization of the mean by removing the seasonal cycle, and the subtraction, in general, of low-frequency energy, are no doubt arbitrary procedures since the governing dynamical equations are nonlinear. In other words, the low-frequency components may be dependent upon the high-frequency components, and *vice versa*. Filtering is therefore justified only so long as the filtered series are not mistaken for the unfiltered series with their great complexity and to the extent that filtering actually enhances the understanding of all the components of the original series,including their mutual interaction.

Both dynamic and synoptic meteorologists have been historically more concerned with the study of higher-frequency weather phenomena rather than with the low-frequency oscillations of the type mentioned above. Hence statistical analysis of time series can aid the development of synoptic and dynamic meteorology by focusing on the phenomena of interest in these branches of meteorology. In this context, we have found that often the filtered time series not containing low-frequency oscillations have autoregressive models of order 2 or more, with complex roots for the corresponding homogeneous difference equations (see also Straus and Halem, 1981). These complex roots are representative of oscillatory mechanisms. Oscillatory mechanisms, such as the wave propagation studied by Rossby (1939, 1945), Parry and Roe (1952) and others, are important conceptual aids in dynamic and synoptic meteorology. Some connections between the oscillatory character of autoregressive processes of order 2 and the weather will be discussed in a future paper.

Acknowledgments

The authors are grateful to Dr. Ivan Kantor, for reviewing the manuscript, and to Fátima Ricco dos Santos who did the typing. They are also grateful to Will Spangler for his help in acquiring the NMC data and to the reviewers for their many helpful comments. This research was partially supported by Financiadora de Estudos e Projetos (FINEP) through Convênio B54/81/042/00/00 - FINEP/CNPq/INPE and Programa Trópico Semi-Árido do CNPq - Pólo Nordeste em 37/38.

References

Blackmon, M.L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, *33*, 1607-1623.

______, J. M. Wallace, N. -C. Lau and S.L. Mullen, 1977: An observational study of the Northern Hemisphere Wintertime circulation. *J. Atmos. Sci.*, *34*, 1040-1053.

Box, G.E.P., and G.M. Jenkins, 1970: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 553 pp.

Childers, D.G., (Editor) 1978: *Modern Spectrum Analysis*. Wiley, New York, 334 pp.

Gandolfo, G., 1971: *Mathematical Methods and Models in Economic Dynamics*. North Holland, Amsterdam, 511 pp.

Gilman, D.L., F.J. Fuglister and J.M. Mitchell, 1963: On the power spectrum of "red noise". *J. Atmos. Sci.*, *20*, 182-184.

Godin, G., 1972: *The Analysis of Tides*. Univ. of Toronto, Toronto, Canada, 77-78.

Haan, C.T., 1977: *Statistical Methods in Hydrology*. Iowa State Univ. Press, Ames. 378 pp.

Hovmöller, E., 1949: The trough-and-ridge diagram. *Tellus*, *1*, 62-66.

Jenkins, G.M., and D.G. Watts, 1968: *Spectral Analysis and Its Applications*. Holden-Day, San Francisco, 525 pp.

Jones, R.H., 1964: Spectral analysis and linear prediction of meteorological time series. *J.Appl. Meteor.*,*3*, 45-52.

______, 1971: Spectrum estimation and time series analysis: A review. Preprints of papers presented at the International Symposium on Probability and Statistics in the Atmospheric Sciences, June 1-4, 1971, Honolulu, Hawaii. Amer. Meteor. Soc., 35-40.

______, and J.P. Kearns, 1976: Fortaleza, Ceará, Brazil rainfall. *J. Appl. Meteor.*, *15*, 307-308.

Julian, P.R., 1971: Investigations of some aspects of non-stationary behavior of synoptic-scale motion systems in the tropics. Preprints of papers presented at the International Symposium on Probability and Statistics in the Atmospheric Sciences.June 1-4, 1971, Honolulu, Hawaii. Amer. Meteor. Soc., 47-52.

Kashyap, R.L., and A.R. Rao, 1976: *Dynamic Stochastic Models from Empirical Data*. Academic Press, New York; p. 47.

Leith, C.E., 1975: The design of a statistical-dynamical climate model and statistical constraints on the predictability of climate. GARP Publication nº 16, WMO, Geneva, Switzerland, 137-141.

Lorenz, E.N., 1965: On the possible reasons for long-period fluctuations of the general circulation. Technical Note nº 66. WMO, Geneva, Switzerland, 203-211.

Papoulis, A., 1962: *The Fourier Integral and Its Application*. McGraw-Hill, New York, 15-24.

Parry, H., and C. Roe, 1952: Record low temperatures in the mid-Atlantic and east central states, October 20-22, 1952. *Mon. Wea. Rev.*, *80*, 195-202.

Parzen, E., 1974: Some recent advances in time series modeling. *IEEE Trans. Autom. Control, AC-19,* 723-730.

Phillips, N., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus, 6,* 273-286.

______, 1956: The general circulation of the atmosphere: a numerical experiment. *Quart. J. Roy. Meteor. Soc., 82,* 123-164.

Rossby, C. -G., and collaborators, 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Marine Res., 2,* 38-55.

______, 1945: On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves. *J. Meteor, 2,* 187-204.

Straus, D.M., and M. Halem, 1981: A stochastic dynamical approach to the study of the natural variability of the climate. *Mon. Wea. Rev., 109,* 407-421.

Tse, F.S., I.E. Morse and R.T. Hinkle, 1963: *Mechanical Vibrations.* Allyn and Bacon, Boston, 34-38.

Ulrych, T.J., and T.N. Bishop, 1975: Maximum entropy spectral analysis and autoregressive decomposition. *Rev. Geophys. Space Phys., 13,* 183-200.

Table Legends

Table 1. Percent variance in the first 4 harmonics of daily data for individual years, as specified in the figure legends.

Table 2. Residual variances for autoregressive models for the time series in which the mean is stationarized by discarding the seasonal cycle. The first line of the table is the variance in the series after the seasonal cycle is discarded. The seasonal cycle is defined as the sum of the first 4 harmonics, with the exception of surface pressure, for which it is the sum of the first 2 harmonics.

Table 3. Positive definite estimates of the autocorrelation function for selected variables.

Table 4. Positive definite estimates of the autocorrelation function for temperature at 850 mb and 41ºN, for various longitudes, and with various modifications of the series. Data are for the 4-year period 1 July 1975 to June 1979.

Table 5. Coefficients of (13), satisfaction of criterion (18), the frequency of free oscillation f_0 and the frequency of the spectral peak f_1, for AR(2). Seasonal cycle filtered out in all instances. The seasonal cycle consists of the sum of the first 4 harmonics in all cases except the surface pressure, for which only the first 2 harmonics were removed.

Table 1. Percent variance in the first 4 harmonics of daily data for individual years, as specified in the figure legends.

Harmonic	Zonal Index	Meridional Index	Surface pressure at		850 mb temperature at		
			Manaus	São Paulo	$41^oN, 80^oW$	$41^oN, 100^oW$	$41^oN, 120^oW$
1	60.3	27.2	39.8	45.1	59.4	55.5	55.9
2	0.1	3.9	10.7	6.7	3.1	3.9	0.8
3	2.3	0.4	0.7	1.5	0.2	0.0	0.6
4	2.1	0.4	2.6	2.2	1.1	0.2	0.9

Table 2. Residual variances for autoregressive models for time series in which the mean is stationarized by discarding the seasonal cycle. The first line of the table is the variance in the series after the seasonal cycle is discarded. The seasonal cycle is defined as the sum of the first 4 harmonics, with the exception of surface pressure, for which it is the sum of the first 2 harmonics.

Order p	Zonal Index m^2s^{-2}	Meridional index m^4s^{-4}	Surface pressure at Manaus mb^2	Surface pressure at São Paulo mb^2	850 mb temperature at 41^oN, 80^oW K^2	850 mb temperature at 41^oN, 100^oW K^2	850 mb temperature at 41^oN, 120^oW K^2
	5.73	4927	1.47	6.59	32.95	37.63	19.41
1	1.99	3400	0.74	3.25	27.80	29.28	10.77
2	1.99	2832	0.72	2.71	27.11	28.24	10.55
3	2.32	2771	0.78	2.75	27.06	27.96	11.10

Table 3. Positive definite estimates of the autocorrelation function for selected variables.

VARIABLE	HARMONICS REMOVED	LAG 1	2	3	4	5	6	7	8	9	10
Meridional Index July 1977 - June 1978	none	0.699	0.338	0.238	0.285	0.323	0.278	0.185	0.149	0.170	0.193
	1-4	0.558	0.029	-0.114	-0.044	0.014	-0.050	-0.184	-0.232	-0.196	-0.157
Station Pressure Manaus 1970	none	0.852	0.702	0.617	0.568	0.549	0.541	0.546	0.550	0.543	0.516
	1-2	0.702	0.404	0.237	0.141	0.103	0.098	0.120	0.136	0.120	0.056
Station Pressure São Paulo 1970	none	0.860	0.665	0.553	0.513	0.512	0.525	0.550	0.559	0.538	0.468
	1-2	0.705	0.301	0.067	-0.012	-0.009	0.026	0.085	0.109	0.074	-0.066
Temperature 850 mb 41°N, 80°W 1976	none	0.774	0.638	0.623	0.638	0.647	0.594	0.566	0.574	0.550	0.549
	1-4	0.393	0.030	0.001	0.060	0.111	-0.019	-0.095	-0.052	-0.090	-0.070
Temperature 850 mb 41°N, 100°W 1976	none	0.799	0.658	0.600	0.585	0.582	0.587	0.549	0.564	0.575	0.569
	1-4	0.479	0.119	-0.020	-0.049	-0.048	-0.024	-0.100	0.054	-0.016	-0.031
Temperature 850 mb 41°N, 120°W 1976	none	0.851	0.717	0.646	0.599	0.557	0.525	0.514	0.532	0.551	0.540
	1-4	0.658	0.349	0.187	0.084	-0.010	-0.076	-0.033	-0.040	0.018	0.005

Table 4. Positive definite estimates of the autocorrelation function for temperature at 850 mb and 41°N, for various longitudes, and with various modifications of the series. Data are for the 4-year period 1 July 1975 to 30 June 1979.

	LAG	LONGITUDE (°W) 130	125	120	115	110	105	100	95	90	85	80	75	70	65
Original, unmodified series	1	0.762	0.839	0.907	0.914	0.879	0.834	0.845	0.877	0.877	0.871	0.855	0.838	0.821	0.804
	2	0.637	0.690	0.807	0.837	0.794	0.717	0.733	0.783	0.782	0.764	0.744	0.725	0.705	0.681
Annual Cycle (fourth harmonic)removed	1	0.585	0.718	0.772	0.705	0.593	0.524	0.545	0.590	0.603	0.611	0.578	0.546	0.516	0.472
	2	0.368	0.458	0.531	0.447	0.309	0.194	0.218	0.282	0.298	0.231	0.259	0.232	0.205	0.143
Estimated seasonal cycle (sum of harmonics 4,8,12 and 16) removed	1	0.569	0.704	0.756	0.699	0.575	0.510	0.535	0.574	0.585	0.593	0.560	0.535	0.511	0.470
	2	0.344	0.430	0.497	0.419	0.279	0.171	0.201	0.255	0.267	0.258	0.227	0.214	0.197	0.139
First 16 harmonics removed	1	0.514	0.672	0.726	0.650	0.529	0.471	0.491	0.531	0.542	0.548	0.512	0.489	0.472	0.440
	2	0.261	0.370	0.436	0.346	0.204	0.105	0.126	0.180	0.191	0.177	0.143	0.136	0.133	0.091

Table 5. Coefficients of (13), satisfaction of criterion (18), the frequency of free oscillation f_o and the frequency of the spectral peak f_1, for AR(2). Seasonal cycle filtered out in all instances. The seasonal cycle consists of the sum of the first 4 harmonics in all cases except the surface pressure, for which only the first 2 harmonics were removed.

	Zonal Index	Meridional Index	Surface pressure at		850 mb temperature at		
			Manaus	São Paulo	41^oN, 80^oW	41^oN, 100^oW	41^oN, 120^oW
β_1	0.805	0.787	0.826	0.995	0.451	0.553	0.754
β_2	-0.018	-0.410	-0.176	-0.405	-0.148	-0.173	-0.147
Inequality (18) satisfied?	No	Yes	Yes	Yes	Yes	Yes	Yes
f_o (cpd)	-	±0.145	±0.028	±0.107	±0.150	±0.134	±0.029
f_1 (cpd)	-	±0.132	-	±0.084	±0.081	±0.057	-

Figure Legends

Figure 1. Positive definite estimate of the autocorrelation function for the zonal index, defined in the text, at 41°N for 1978, with the seasonal cycle retained.

Figure 2. Positive definite estimate of the autocorrelation function for the zonal index at 41°N for 1978, with the seasonal cycle, assumed to be the sum of the first 4 harmonics, discarded.

Figure 3. Smoothed spectral density estimates, for the data in Fig. 2, using the Parzen window (continuous line), and the corresponding first order autoregressive model (dashed line). The bandwidth and 80% confidence interval are indicated.

Figure 4. Smoothed spectral density estimates for the meridional index, defined in the text, for the year 1 July 1977 to 30 June 1978, with the seasonal cycle, consisting of the first 4 harmonics, removed. The estimates computed using the Parzen lag window are shown by the continuous line. The first and second order autoregressive models are represented by the dotted and dashed lines, respectively. The banddwidth and the 80% confidence interval are also shown.

Figure 5. As in Fig. 4, but the smoothed spectral density estimates are for station pressure at a) Manaus and b) São Paulo, with just the first 2 harmonics deleted in each instance.

Figure 6. As in Fig. 4, but the smoothed spectral density estimates are for 850 mb temperature in 1976 at 41°N and a) 80°W, b) 100°W and c) 120°W, with the first 4 harmonics removed in each case.

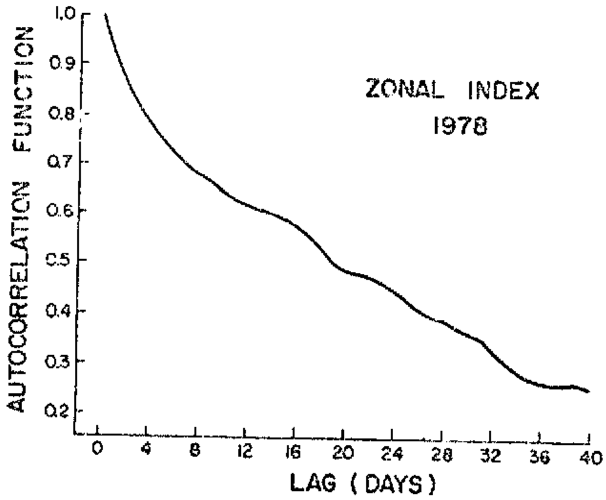
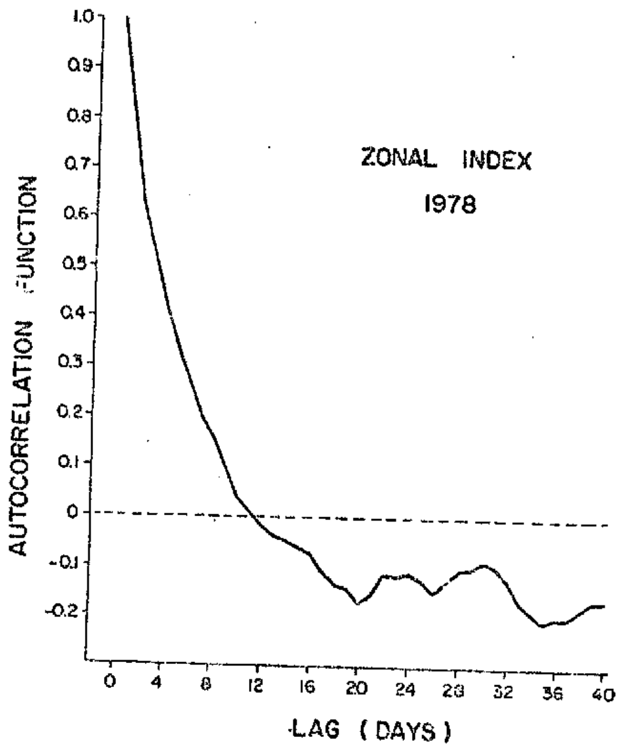

ZONAL INDEX
1978
AUTOCORRELATION FUNCTION
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0 4 8 12 16 20 24 28 32 36 40
LAG (DAYS)

Fig. 1



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
AUTOCORRELATION FUNCTION
ZONAL INDEX
1978
0
4
8
12
16
20
24
28
32
36
40
LAG (DAYS)

Fig. 2

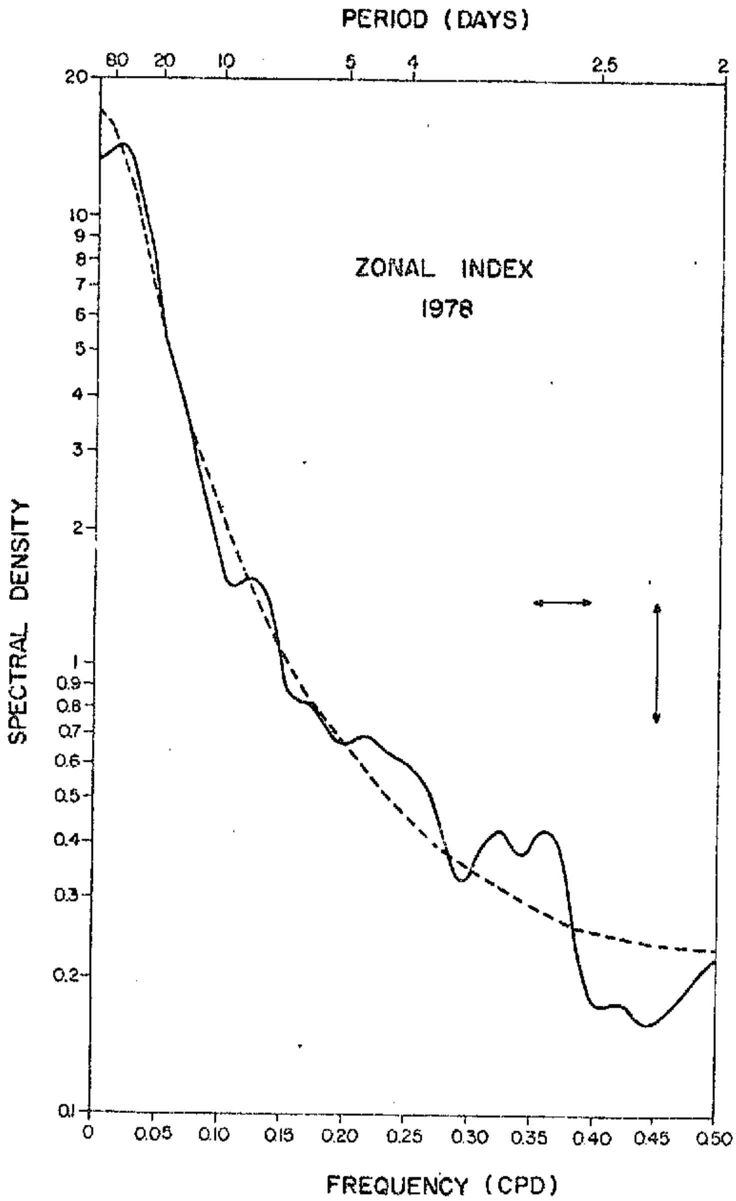

PERIOD (DAYS)
80 20 10 5 4 2.5 2
ZONAL INDEX
1978
SPECTRAL DENSITY
20 10 9 8 7 6 5 4 3 2 1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1
0 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50
FREQUENCY (CPD)

Fig. 3

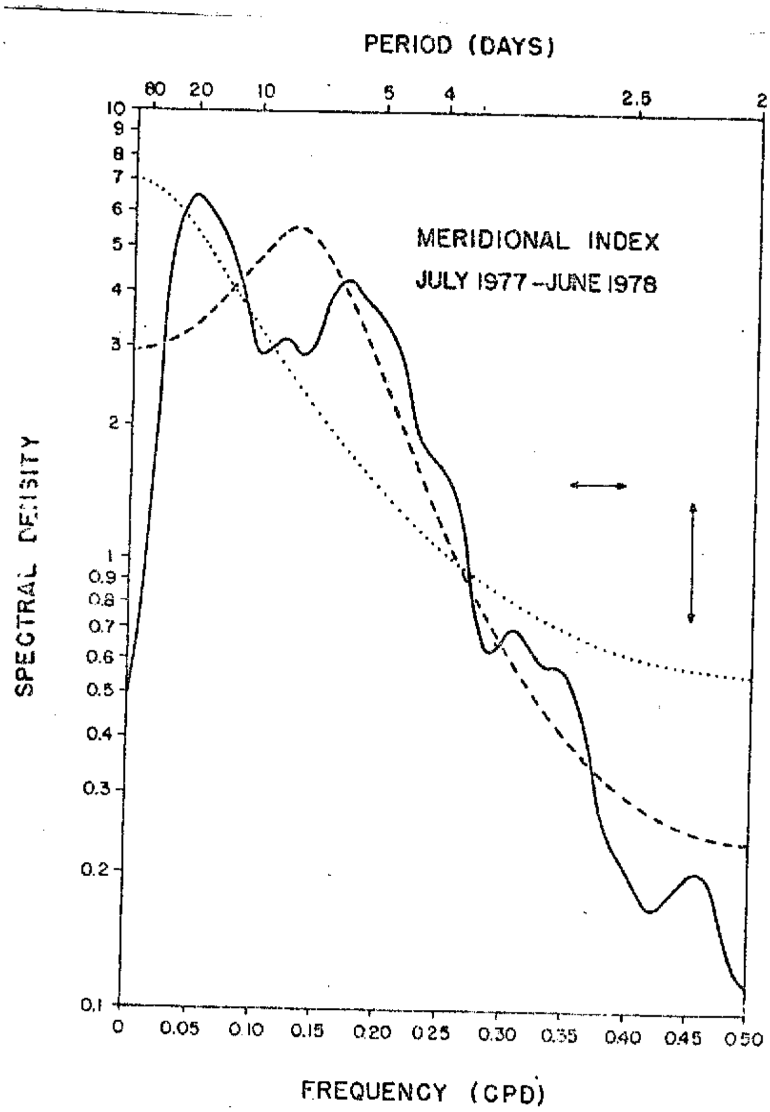
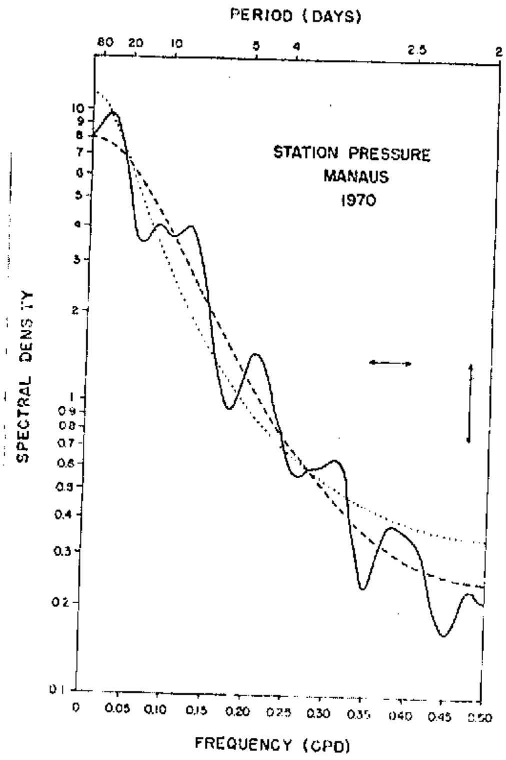
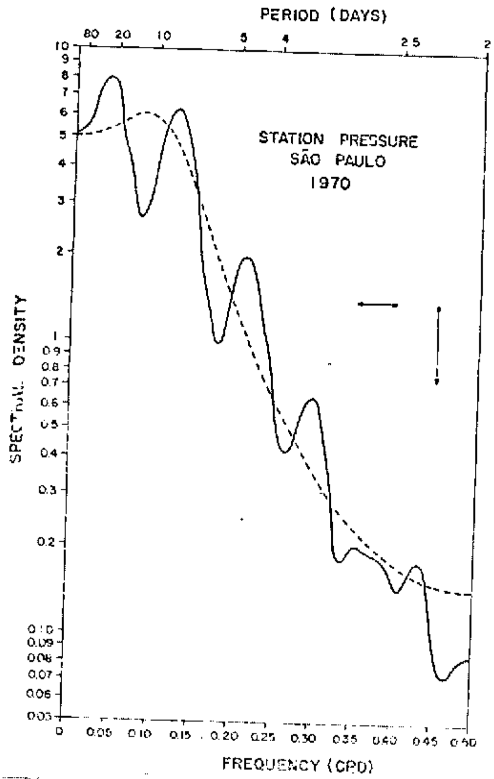

PERIOD (DAYS)
60 20 10 5 4 2.5 2
MERIDIONAL INDEX
JULY 1977-JUNE 1978
SPECTRAL DENSITY
10 9 8 7 6 5 4 3 2 1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1
0 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50
FREQUENCY (CPD)

Fig. 4

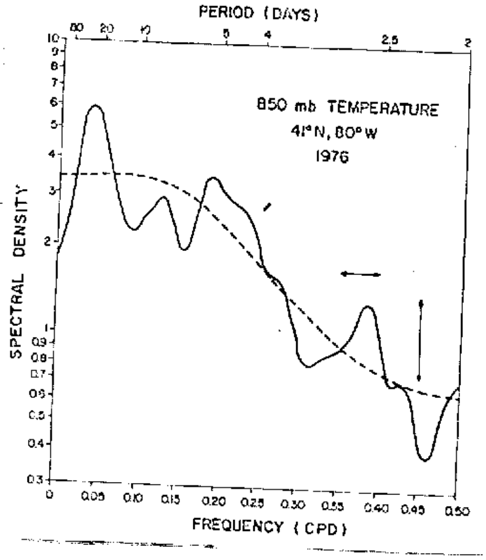

PERIOD (DAYS)
STATION PRESSURE
MANAUS
1970
SPECTRAL DENSITY
FREQUENCY (CPD)

(a)

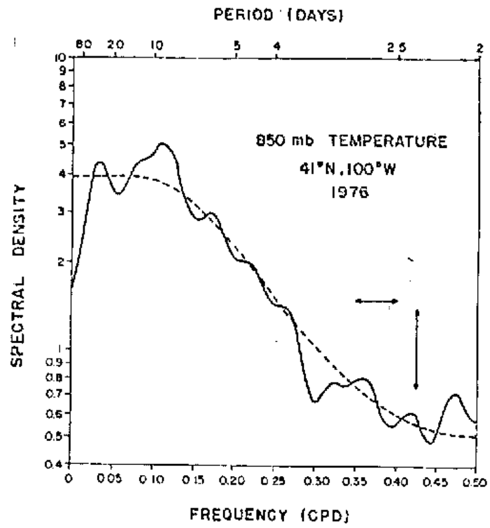

PERIOD (DAYS)
STATION PRESSURE
SÃO PAULO
1970
FREQUENCY (CPD)

(b)

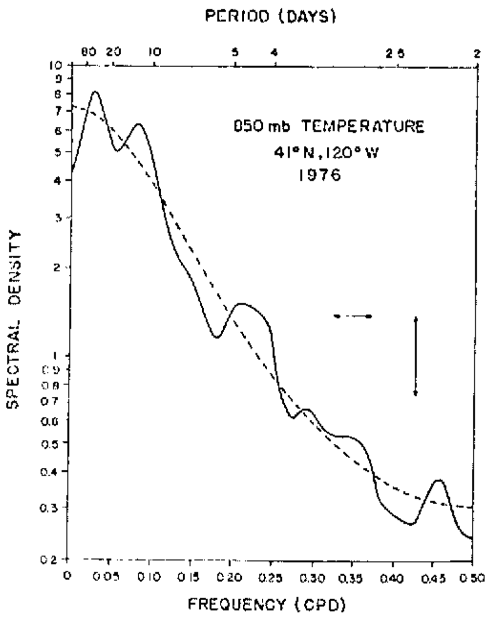
Fig. 5


PERIOD (DAYS)
850 mb TEMPERATURE
41°N, 80°W
1976
SPECTRAL DENSITY
FREQUENCY (CPD)

(a)


PERIOD (DAYS)
850 mb TEMPERATURE
41°N, 100°W
1976
SPECTRAL DENSITY
FREQUENCY (CPD)

(b)


PERIOD (DAYS)
850 mb TEMPERATURE
41°N, 120°W
1976
SPECTRAL DENSITY
FREQUENCY (CPD)

(c)

Fig. 6