Some statistical significance tests useful in spectral analysis of meteorological time series

F.J. Soley*

Resumen

Algunos ensayos estadísticos útiles en análisis espectral se aplican a series temporales meteorológicas provenientes de varias estaciones en el Oeste del Caribe. Los estimadores de la densidad espectral de potencia utilizados en este trabajo son el periodograma, el estimador de Blackman y Tukey, y el modelo autoregresivo-media móvil. En el caso del último estimador se utilizan varios métodos para estimar objetivamente el orden del modelo autoregresivo-media móvil. En la mayoría de las series analizadas se encuentra que el modelo óptimo es un modelo autoregresivo puro de bajo orden. También se demuestra que la distribución beta incompleta es la distribución de probabilidad exacta del periodograma integrado de ruido blanco que se utiliza en uno de los ensayos estadísticos. Para ilustrar la aplicación de ensayos estadísticos a series filtradas, se filtra una serie temporal con un pico espectral aparente con un filtro de paso de banda, y la potencia de salida se contrasta con los dos posibles modelos estocásticos.

Abstract

Some statistical significance tests which are useful in spectral analysis are applied to meteorological time series from several stations in the Western Caribbean. The estimates of the power spectral density used in this work are the periodogram, the Blakcman-Tukey and the autoregressive-moving average. In the last case, several methods of objectively estimating the order of the processes are used and it is found that in most cases the optimum models correspond to pure autoregressive processes of low order. It is shown that the exact probability distribution of the integrated periodogram used in a

* Centro de Investigaciones Geofísicas y Escuela de Física. Universidad de Costa Rica, San José, Costa Rica.

white noise test is the incomplete beta function. A time series with an apparent spectral peak is filtered with a band pass filter and the power of the output tested against two possible stochastic models.

Introduction

During a study of synoptic wave-like disturbances in the tropical regions using spectral analysis, it was recognized the importance of having available objective criteria to evaluate the statistical significance of spectral features, such as, for example, peaks, troughs and red or blue noise type behavior. The large variability of some of the power spectral density estimates can easily mislead the user in believing that an spurious peak is real, or to regard two spectra as due to different processes when in reality the differences could be the natural ones arising in the realizations of the same stochastic model. Even though statistical significance tests are very useful as guides, they should be complemented in the final analysis with the known physical properties of the variable involved.

For many years spectral analysis has been used succesfully in the analysis of meteorogical time series. Jones (1972) has listed some of its main uses:

- to locate the dominant scales of motion,
- to investigate the structure and dynamics of the dominant wave modes,
- the possible prediction of future behavior of the time series,
- verification of dynamical theories, and
- to locate spectral shapes obeying power laws.

It is also well known that great care hast to be exercised in the practical application of spectral analysis and in the interpretation of the results (Jones, 1972; Gutowski et al, 1978). It is the purpose of this work to present some statistical significance tests and objective criteria which can be useful in spectral analysis, and to illustrate their applications with meteorological time series from several stations in the Western Caribbean. The need of building a complete statistical picture of the data will be stressed.

Data and methods

The geopotential, the meridional and zonal wind components at the tropospheric levels of 850, 700 and 300 hPa were analyzed for periods of several months in the second half of the year and for different years and stations. The stations and years used were: Swan Island (17°24'N, 83°56'W) and San Andres Island (12°35'N, 81°43'W) for the years (1970, 1971 and 1972; Santamaria (10°00'N, 84°13'W) and Howard (8°59'N, 79°33'W) for the years 1972 and 1979. Each of the time series was inspected for both gaps and data discontinuities. Subjective synoptic analysis of data in the region and simple linear time

interpolation were applied to missing data. Howard station showed the largest percentage of missing data (7%) for the periods analyzed.

A comprehensive review of the available methods for estimating the power spectral density can be found in Kay and Marple (1981). In this work we will adhere closely to their definitions and conventions, unless otherwise indicated.

Periodogram

It is calculated using the Fast Fourier Transform (FFT) algorithm developed by Sande and the prunning procedure of Markel (Robinson, 1983). It must be remembered that the periodogram spectral estimate is not scaled properly as a power spectral density, since it is the value of the peak, rather than the area under the peak that is equal to power (Kay and Marple, 1981). The data sometimes has to be zero padded to comply with a total length equal to a power of two. Zero padding interpolates additional values between the independent frequency components of then non-zero padded transform and lowers the true average power of the time series. The first effect causes no trouble if the statistical significance tests are carried out with the true number of independent frequency components. The second effect is taken into account by scaling the power spectra so that the sum of its terms equals the average power of the time series according to Parseval's Theorem (Cooley et al, 1969).

The statistical significance of this estimate can be tested with the known relation between the periodogram of white noise and the chi-square distribution (Olberg, 1982; Murphy and Katz, 1985). Let P_J be the value of the periodogram at frequency $f_J = J/N\tau$, where N is the order of the transform and τ the sampling time. Then:

i) $P_J/(\sigma^2/2N)$ is distributed as a chi-square distribution of order 2 for $J = 1,2,3\ldots, (N/2)-1$, and with σ^2 the average power.
ii) $P_0=0$, since in our case the mean is subtracted from the data.
iii) for $J = N/2$, $P_J/(\sigma^2/N)$ is distributed as a chi-square distribution of order 1.

White noise test

A time series can be tested against white noise by integrating the periodogram (Jenkins and Watts, 1968; Jones and Hearns, 1976). The significance levels usually used are calculated assuming a normal distribution and are independent of frequency. This independence implies that negative power values are possible near zero frequency and that near the Nyquist frequency the integrated power can be larger than the observed total variance. In practice, the end points of the integrated spectrum are fixed because the Nyquist frequency term is one for normalized spectra and the zero frequency or DC term is zero since usually the mean is subtracted from the series. An expression for the probability distribution function of the integrated spectrum, which complies with fixed end points and is based on the chi-square statistics of the periodogram estimate, is derived as follows

Let $Z_m = 2 * \sum_{j=1}^{m} P_j$, for m=1,2,3,......, (N/2)−1, then the

normalized variable $w_m = N/\sigma^2 Z_m$, follows a chi-square

distribution with 2m degrees of freedom, since it is the sum of m stochastic variables distributed as chi square of order two. Similarly, the total variance σ^2 is distributed as chi-square distribution with 2N−1 degrees of freedom. Therefore we are interested in the probability that w_m/σ^2 obtains a value less than x under the condition that the variance obtained a value s^2, that is, $P\{w_m/\sigma^2 < x \mid \sigma^2=s^2\}$, which is the incomplete beta function $I_x(a,b)$ with a=m and b= N−m−1/2 (Abramowitz and Stegun, 1970).

Blackman-Tukey estimate

The Blackman-Tukey (B-T) estimate of the power spectral density (Blackman and Tukey, 1959) is calculated using the biased autocorrelation estimate and the window proposed by Papoulis (1973), which has a slightly smaller variance than Parzen's window (Jenkins and Watts, 1968). For this window, the degrees of freedom are 3.73 N/M and the badnwidth 1.87/M, where N is the length of the time series and M the length of the autocorrelation. If N-1 samples of the autocorrelation are computed, then the unwindowed B-T estimate and the periodogram yield identical numerical results (Kay and Marple, 1981).

The statistical significance of the B-T estimate P(f), against the value $P_m(f)$ predicted by a given model, can be calculated computing the ratio.

$$P(f)/P_m(f) > CC^2/n, \quad (1)$$

where CC is the significance level of a chi-square distribution of n degrees of freedom (Olberg, 1982).

Olberg uses Blackman-Tukey's estimate of the number of degrees of freedom, that is, n=2*(N/M − 1/3), which differs from the degrees of freedom of the window proposed by Papoulis. The statistical significance level for a peak in the B-T estimate should be somewhat higher or about the same than the significance level for the same peak in the periodogram estimate, since the B-T estimate can be considered as a smoothing of the periodogram. It was found that the statistical significance tests are in agreement if one uses the degrees of freedom suggested by Papoulis, whereas the estimates with Blackman-Tukey's proposed degrees of freedom tend to underestimate the significance levels.

Autoregressive-Moving average power spectral density estimate

If the data can be represented by an autoregressive-moving average (ARMA(p,q)) model, the corresponding power spectral density estimate is:

$$P(f) = \frac{\sigma^2\tau[1+\sum b_k \exp(-j2\mu fk\tau)]^2}{[1+\sum a_k \exp(-j2\mu fk\tau)]^2} \quad (2)$$

where a_k are the p autoregressive coefficients, b_k the q moving average coefficients, σ^2 the variance of the innovation process and τ the sampling time interval (Kay and Marple, 1981).

In practice, diagnostic checks must be carried out to see if the more general autoregressive-integrated-moving average ARIMA (p,d,q) process is necessary to model the data (Box and Jenkins, 1970). A first estimate of the model order is obtained from inspection of the autocorrelation and partial autocorrelation functions. A final decision can be made inspecting the Bayesian Information Criterion (BIC) (Katz and Skaggs, 1981), the Minimum Akaike Information Criterion (MAIC) (Ozaki, 1971), and the Generalized Partial Autocorrelation function (GPAC) (Woodward and Gray, 1981). The first two are based on the residual variance, which decreases with increasing model complexity, and a function, which is different for the two criteria, that penalizes the increase in the number of parameters in the model. Thus, the optimum number of parameters corresponds to those where the criteria is a minimum. The third criteria is based in the generalized partial autocorrelation function whose values are arranged in matrix form with rows labeled by the MA order, starting with order 0, and columns labeled by the AR order starting with 1. One way of identifying the model order is by locating a column with nearly constant values, accompanied by a row of entries near zero. The label of the column and row are the values of p and q, respectively. The data used in this work has been found to correspond to ARIMA (0,0,p) processes, that is, pure autoregressive processes AR(p). The details of parameter estimation for AR(p) processes are given below.

Autoregressive power spectral density estimate

Of the several methods for estimating AR parameters, the forward-backward linear prediction version (Robinson, 1983) was chosen because of its higher resolution, lack of observed line splitting, reduced bias in the power estimates and absence of sidelobes (Kay and Marple, 1981). An objective estimate of the model order can be obtained from the Final Prediction Error coefficients (FPE), Akaike's Information Criterion (AIC), and with Parzen's Criteria Autoregressive Transfer function (CAT) (Kay and Marple, 1981). For the data used in this paper the three methods agree in the order selected, except for very few exceptions. Once the order p and the AR coefficients a_k are determined, the spectra is calculated with Ec. 2, setting all b_k coeficients to zero.

Filtered time series

In many applications it is useful to filter the data to enhance certain spectral feature and/or to reduce noise. Olberg (1982) has derived expressions to test the significance of the ratio of the power of the filtered data and the power of a proposed model series filtered in the same way. For example, when looking for an spectral peak, a band pass filter may be applied to the data and the power of the filtered series tested against a white noise model. The test gives the probability that the power observed may be due to an spurious peak within the filter's band pass.

Results

Choosing the appropriate model

Fig. 1 shows the autocorrelation and partial autocorrelation coefficients for the geopotential at the 850, 700 and 300 hPa levels at Howard Base during the period June 1 to August 31, 1979. In all three cases the autocorrelation function does not dampen out, whereas the partial autocorrelation coefficients are close to zero after the third or fourth. From this information one chooses tentatively an AR(3) process for the 850 and 300 hPa levels and an AR(4) for the 700 hPa level, although in this last case, the selection is dubious because the fourth coefficient is between one and two standard deviations. On inspection of the GPAC functions (Tables 1,2,3), it can be concluded that all three series can be modeled with an order three autoregressive process, although the near constant behaviour is lost at high q values. Table 4 shows the BIC and the MAIC for the 850 and 700 hPa levels, which confirms that an AR(3) process is the right choice, except in the case of the MAIC for the 700 hPa level that selects an order 4 process. Notice that the BIC has a more pronounced minimum than the MAIC. All the criteria above, point to pure autoregressive processes, and only the order of the 700 hPa level is doubtfull. The FPE, AIC and CAT functions, which are applicable to AR processes, predicted each one of them consistently for the 850, 700 and 300 hPa levels, orders of 3,4 and 3, respectively, which can again be traced to the fact that the 700 hPa level has the fourth partial autocorrelation coefficient near the significance threshold boundary. Fig. 1 also compares the Blackman-Tukey power spectral density estimate with the autoregressive estimates. The differences in the spectra are not statistically significant. For example, the peak near 0.25 day^{-1} for the B-T estimate of the 850 hPa level has a significance level of 0.15 when compared with the power density predicted by the AR(3)

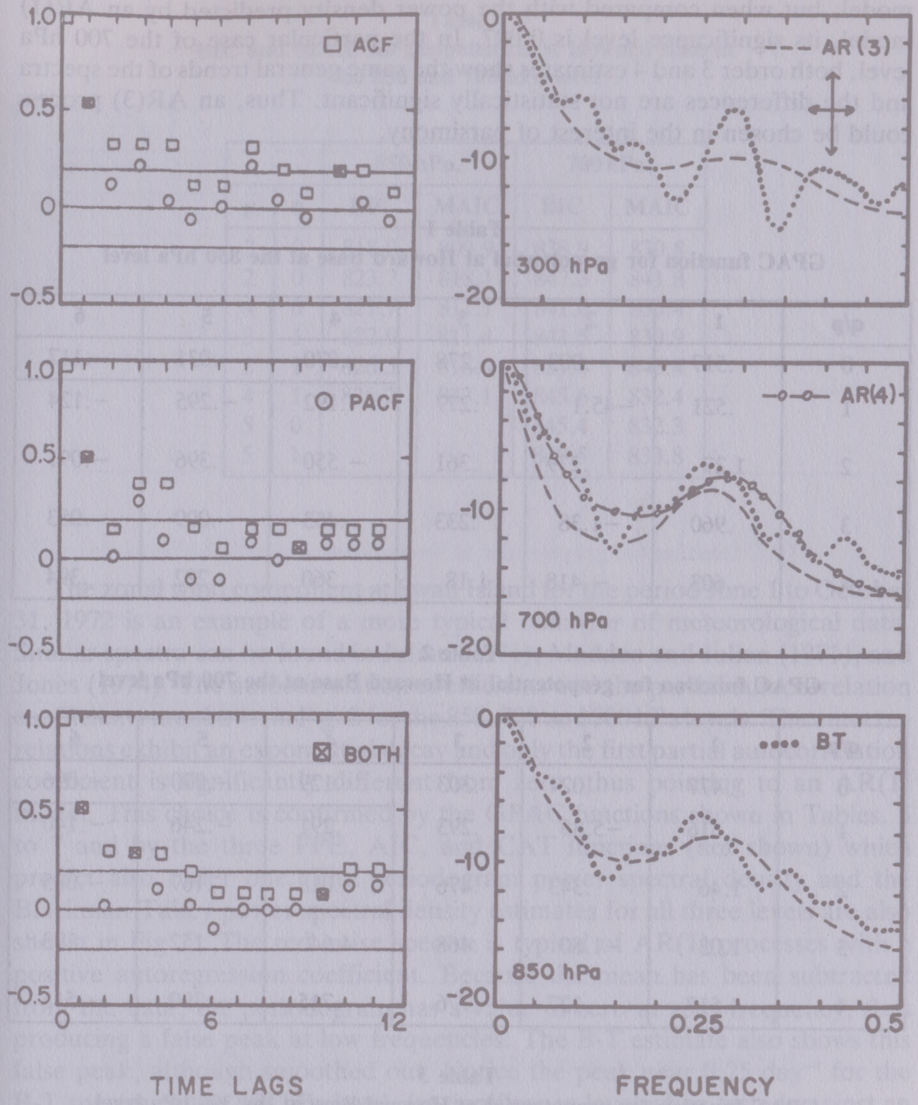

Figure 1. The autocorrelation (ACF) and partial autocorrelation (PACF) functions for the geopotential at the 850, 700 and 300 hPa levels at Howard Base are shown on the left. The Blackman-Tukey (BT) and autoregressive (AR) power spectral density estimates at normalized frequency are shown on the right. The sampling interval is 1 day. The value corresponding to two standard deviations for the PACF is indicated by the horizontal lines. The cross indicates the 0.01 significance level and the bandwidth of the BT estimate.

model, but when compared with the power density predicted by an AR(1) model, its significance level is 0.007. In the particular case of the 700 hPa level, both order 3 and 4 estimates show the same general trends of the spectra and the differences are not statistically significant. Thus, an AR(3) process could be chosen in the interest of parsimony.

Table 1
GPAC function for geopotential at Howard Base at the 850 hPa level

q/p	1	2	3	4	5	6
0	.517	.003	.278	.070	−.031	−.117
1	.521	−45.1	.277	.192	−.295	−.124
2	1.27	.279	.361	−.550	.396	−.091
3	.960	−1.38	.233	.453	.099	−.053
4	.603	−.418	1.18	.360	.282	.394

Table 2
GPAC function for geopotential at Howard Base at the 700 hPa level

q/p	1	2	3	4	5	6
0	.477	.024	.303	.129	−.070	−.096
1	.516	−5.85	.293	.291	−.246	−.140
2	1.46	.343	.476	−.018	.167	−.085
3	1.03	−1.80	.468	4.64	.152	.193
4	.517	−.127	.776	.715	−.390	−.515

Table 3
GPAC function for geopotential at Howard Base at the 300 hPa level

q/p	1	2	3	4	5	6
0	.499	.122	.187	.031	−.100	.041
1	.682	−.630	.167	.627	−.087	.604
2	−1.02	.485	.418	.185	.468	.131
3	.786	−1.07	.741	−.556	.138	−.401

Table 4
BIC and MAIC for geopotential at Howard Base at the 850 and 700 hPa level

		850 hPa.		700 hPa.	
p	q	BIC	MAIC	BIC	MAIC
3	0	818.0	809.9	838.9	830.8
2	0	823.7	818.1	847.3	841.8
4	0	821.7	811.1	841.0	830.4
3	1	822.0	811.4	841.5	830.9
2	1	826.5	818.4	851.4	843.4
4	1	826.2	813.1	845.6	832.4
5	0			845.4	832.3
5	1			849.5	833.8

The zonal wind component at Swan Island for the period June 1 to October 31, 1972 is an example of a more typical behavior of meteorological data. Similar spectra can be found in Julian (1971), Madden and Julian (1971), and Jones (1974). The autocorrelation coefficients and the partial autocorrelation coefficients are shown in Fig. 2 for the 850, 700 and 300 hPa levels. The autocorrelations exhibit an exponential decay and only the first partial autocorrelation coefficient is significantly different from zero, thus pointing to an AR(1) model. This choice is confirmed by the GPAC functions shown in Tables. 5 to 7 and by the three FPE, AIC, and CAT functions (not shown) which predict also order one. The periodogram power spectral density and the Blackman-Tukey power spectral density estimates for all three levels are also shown in Fig. 2. The red noise spectra is typical of AR(1) processes with a positive autoregression coefficient. Because the mean has been subtracted from the data, the periodogram has a value of zero at zero frequency, thus producing a false peak at low frequencies. The B-T estimate also shows this false peak, although smoothed out. Notice the peak near 0.25 day^{-1} for the B-T spectra of the 850 hPa level. Its significance level when tested against an AR(1) model gives .0193. Since all the evidence points out that the AR(1) model is correct, this peak can be considered spurious, in spite of its high significance level. Similarly, the bumps at midfrequencies for the 300 hPa level are all spurious with significance levels between .2 and .1.

White-noise test

The geopotential for the period from June 1 to Aug. 31, 1979 at Santamaria Airport in Costa Rica, also presented the AR(1) behaviour in all three levels.

Table 5

GPAC function for geopotential at Swan Island at the 850 hPa level

q/p	1	2	3	4	5	6
0	.803	.073	.052	−.054	−.057	−.006
1	.836	−.050	.128	−.110	−.051	−.057
2	.860	1.17	.747	−.049	−.037	.865
3	.820	−.647	.372	−.367	−.331	.220
4	.784	.085	.053	1.80	1.56	.065
5	.789	−.473	−2.40	1.53	1.12	−.849

Table 6

GPAC function for geopotential at Swan Island at the 700 hPa level

q/p	1	2	3	4	5	6
0	.786	.107	−.015	.030	−.087	.061
1	.838	.216	.201	−.013	−.066	.098
2	.821	1.00	.402	−1.47	−.070	.128
3	.846	2.20	−3.28	.695	.119	.063
4	.767	.704	.994	1.52	−.870	−.014
5	.853	−.033	−.415	.848	−1.07	−6.19

Table 7

GPAC function for geopotential at Swan Island at the 300 hPa level

q/p	1	2	3	4	5	6
0	.604	.053	−.038	.077	−.059	.037
1	.660	.485	.059	.037	−.015	.035
2	.592	1.14	−.732	.049	.046	.077
3	.790	.629	.294	−.161	−.276	.198
4	.524	.231	−1.48	2.51	−1.12	−12.2

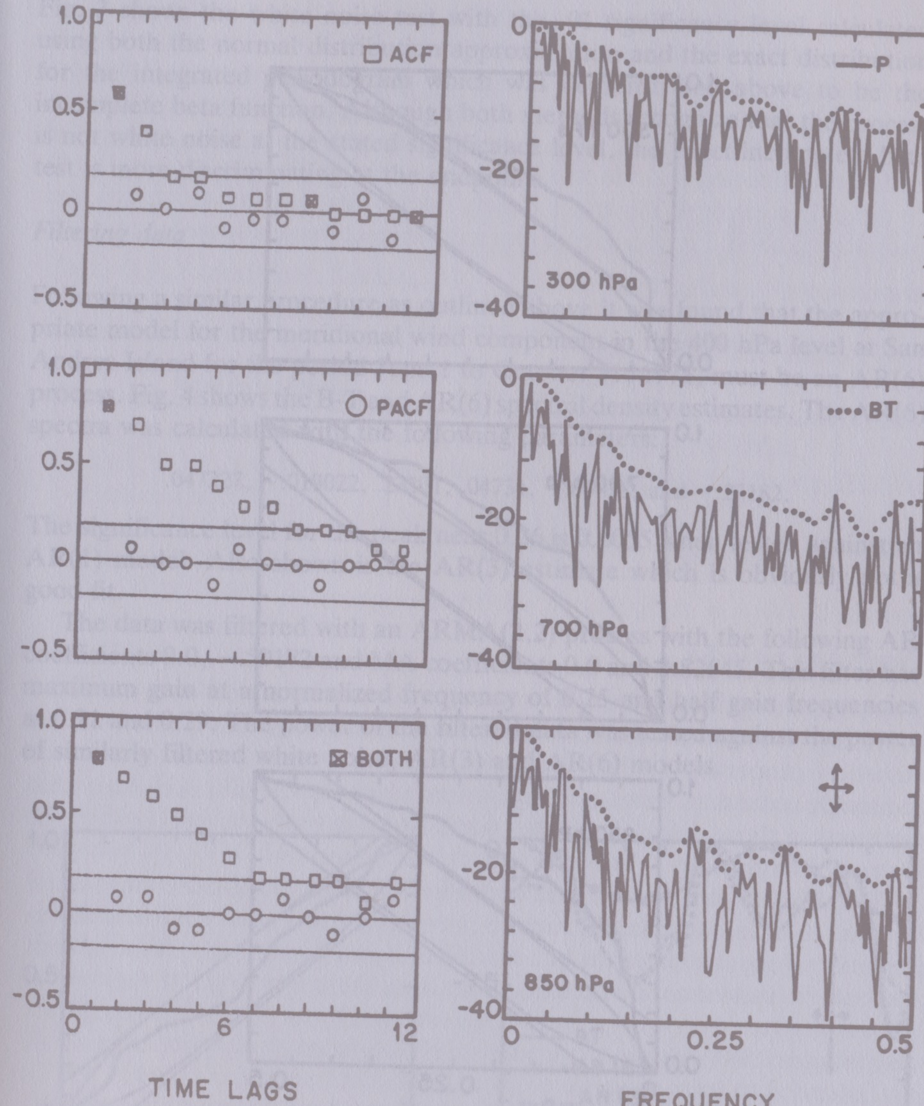

Figure 2. The autocorrelation (ACF) and partial autocorrelation (PACF) functions for the zonal wind component at the 850, 700 and 300 hPa levels at Swan Island are shown on the left. The periodogram (P) and Blackman-Tukey (BT) power spectral density estimates at normalized frequency are shown on the right. The sampling interval is 1 day. The value corresponding to two standard deviations for the PACF is indicated by the horizontal lines. The cross indicates the 0.01 significance level and the bandwidth of the BT estimate.

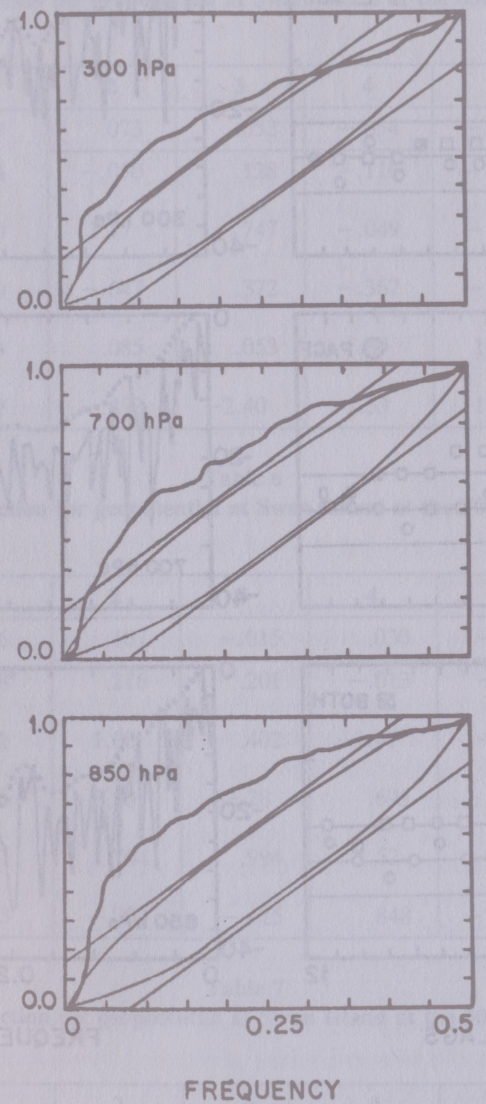

Figure 3. White noise test for the geopotential at the 850, 700 and 300 hPa levels at Santamaria Airport. The 0.01 significance level is shown calculated according to the normal distribution approximation (straight lines) and according to the incomplete beta function (curved lines that pass through the fixed endpoints).

Fig. 3 shows the white noise test with the .01 significance level calculated using both the normal distribution approximation and the exact distribution for the integrated periodogram which was demonstrated above to be the incomplete beta function. Although both methods recognize that the process is not white noise at the stated significance level, the exact incomplete beta test is more discriminating at the endpoints.

Filtering data

Following a similar procedure as outlined above it was found that the appropriate model for the meridional wind component in the 400 hPa level at San Andres Island for the period June 1 to October 31, 1971, must be an AR(6) process. Fig. 4 shows the B-T and AR(6) spectral density estimates. The AR(6) spectra was calculated with the following parameters:

.047227, −.019022, .23101, .04731, −.05628, and −.27182.

The significance level for the peak near 0.26 is 0.0055 when tested against an AR(1) model. Also shown is the AR(3) estimate which is obviously not a good fit.

The data was filtered with an ARMA(2,2) process with the following AR coefficients 0.0, −.59172 and MA coefficients 0.0 and 0.82645. This filter has maximum gain at a normalized frequency of 0.25 and half gain frequencies at 0.21 and 0.29. The power of the filtered data was tested against the power of similarly filtered white noise, AR(3) and AR(6) models.

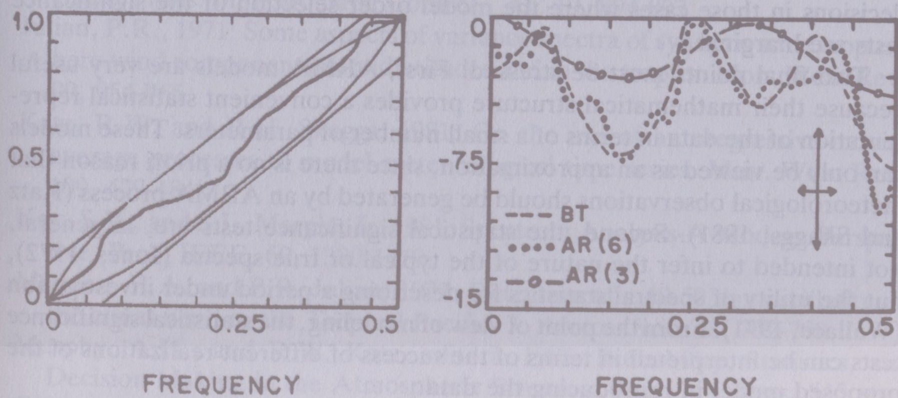

Figure 4. White noise test at the 0.01 significance level (left) and Blackman-Tukey (BT) and autoregressive (AR) power spectral density estimates (right) of the meridional wind component at the 700 hPa level at San Andres Island. The sampling interval is 1 day. The cross indicates the 0.01 significance level and the bandwidth of the BT estimate.

Table 8 shows that output power of the filtered AR(6) model is closest to the power of the filtered data, but neither the AR(3) or white noise models are significantly different at the .05 level, since their significance limits are below 1.96. An inspection of the white noise test (Fig. 4), shows that the data is only marginally different from a white noise process, which explains the results from filtering and the small autoregressive coefficients required to fit the data. This example illustrates clearly that in this case the apparently real spectral peak should be considered spurious, unless other evidence is provided.

Table 8
Power ratio test

	Power ratio	Significance limit
AR(3)	1.2212	1.1293
AR(6)	.9883	.0478
White noise	1.2518	1.2934

Concluding remarks

Complementing the traditional spectral analysis techniques with ARIMA modeling, fitting and checking, provides a fairly complete picture of the statistical properties of the time series being analyzed. The careful use of seemingly redundant tests and checks turns to be very fruitful in making final decisions in those cases where the model order selection or the significance tests are marginal.

Two final points must be stressed. First, ARMA models are very useful because their mathematical structure provides a convenient statistical representation of the data in terms of a small number of parameters. These models can only be viewed as an approximation, since there is no a priori reason that meteorological observations should be generated by an ARMA process (Katz and Skaggs, 1981). Second, the statistical significance tests are, in general, not intended to infer the nature of the typical or true spectra (Jones, 1972), but the utility of spectral statistics for describing a period under investigation (Wallace, 1971). From the point of view of modeling, the statistical significance tests can be interpreted in terms of the success of different realizations of the proposed model in reproducing the data.

Acknowledgements

The author is indebted to Jorge Amador for his helpful comments on the manuscript and to Roger Sipson for part of the programming work. Thanks are also due to Mario Fernández for drawing the figures.

References

Abramowitz, M., and I.A. Stegun, editors, 1970: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards, Applied Mathematics Series, 55. U.S. Government Printing Office, 1046 pp.

Blackman, R.B., and J.W. Tukey, 1959: The Measurement of Power Spectra from the Point of View of Communications Engineering. Dover, 190 pp.

Box, G.E.P. and G.M. Jenkins, 1970: Time Series Analysis: Forecasting and Control. Holden-Day, 575 pp.

Cooley, J.W., D.A.W. Lewis and P.D. Welch, 1969: The finite Fourier transform. IEEE Trans. Audio Electroacoust., AU-17, 77-85.

Gutowski, P.R., E.A. Robinson and S. Treitel, 1978: Spectral estimation: fact or fiction. IEEE Trans. Geosc. Electron., GE-16, 80-84.

Jenkins, G.M., and D.G. Watts, 1968: Spectral Analysis and Its Applications. Holden-Day, 543 pp.

Jones, R.H., and J.P. Hearns, 1976: Fortaleza, Ceará, Brazil Rainfall. J. Appl. Meteor., 15, 307-308.

Jones, R.H., 1965: A reappraisal of the periodogram in spectral analysis. Technometrics, 7, 531-542.

Jones, R.H., 1972: The use and abuse of spectral analysis in tropical meteorology. In "Dynamics of the tropical atmosphere. Notes from a colloquium: summer 1972". National Center for Atmospheric Research. Boulder, Colorado, 587 pages.

Jones, R.H., 1974: Identification and autoregressive spectrum estimation. IEEE Trans. on Automatic Control., AC-19, 894-897.

Julian, P.R., 1971: Some aspects of variance spectra of synoptic scale tropospphere wind components in midlatitudes and in the tropics. Mon. Wea. Rev., 99, 954-965.

Katz, R.W. and R.H. Skaggs, 1981: On the use of autoregressive-moving average processes to model meteorological time series. Mon. Wea. Rev., 109, 479-484.

Kay, S.V., and S.L. Marple, Jr. 1981: Spectrum Analysis-A Modern Perspective. Proc. IEEE, 69, 1380-1419.

Madden, R.A. and P.R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the Tropical Pacific. J. Atmos. Sci., 28, 702-708.

Murphy, A.H., and R.W. Katz, editors, 1985: Probability, Statistics, and Decision Making in the Atmospheric Sciences. Westview Press, 545 pp.

Olberg, M., 1982: Zür statistischen Beurteilung von Analyseergebnissen geophysikalisch-meteorologischer Beobachtungsreihen. Gerlands Beitr. Geophysik, 91, 281-290.

Ozaki, T., 1977: On the order determination of ARIMA models. Appl. Statist., 26, 290-301.

Papoulis, A., 1973: Minimum-bias windows for high resolution spectral estimates. IEEE Trans. on Inform. Theory, IT 19, 9-12.

Robinson, E.A., 1983: Multichannel Time Series Analysis with Digital Computer Programs. Goose Pond Press, 454 pp.

Woodward, W.A. and H.L. Gray, 1981: On the relationship between the S array and the Box-Jenkins method of ARMA model identification. Jour. Amer. Stat. Assoc., 76, 579-587.