

ACERCA DEL ANALISIS ESPECTRAL POR EL METODO DE MAXIMA ENTROPIA.

Humberto Fuenzalida

Departamento de Geofísica. Universidad de Chile.
Casilla 2777. Santiago-Chile.

RESUMEN

En este trabajo se presentan los resultados de algunas aplicaciones del Método de Máxima Entropía para estimar la densidad espectral de series cronológicas sintéticas y observadas. Las bases teóricas del método son desarrolladas en términos del filtro predictor mostrándose su relación con modelos autorregresivos. Se comentan las determinaciones de la forma y potencia de las líneas espectrales y una técnica aproximada para estimar el grado de significancia de cada una de ellas. Diversos criterios para fijar el orden del filtro predictor son consideradas.

Se presentan los resultados del análisis de series de ruido blanco, series autorregresivas, series de números de manchas solares y de pluviometría anual en estaciones de la zona semiárida de Chile. Se evidencia una modesta dependencia del orden del filtro en el espectro de series modelables por procesos autorregresivos, no así en series de ruido blanco ni en las series pluviométricas. En este último caso los criterios de orden usados no identifican ningún valor particular aunque el de significancia es satisfecho para algunas líneas con frecuencias asociables al ciclo cuasibienal.

En general, parece recomendable hacer un barrido de órdenes, ya que el grado de significancia de las líneas espectrales se optimiza en un valor intermedio que no es señalado por los criterios conocidos. Las series de manchas solares aparecen modelables por un proceso autorregresivo de tercer orden.

ABSTRACT

ON THE SPECTRAL ANALYSIS BY THE MAXIMUM ENTROPY METHOD.

Results of spectral analysis by the Maximum Entropy Method to synthetic and real time series are presented. The theoretical background of the method is developed along the lines of the prediction filter for time series, stressing its relation to the autoregressive model. Techniques which provide power, shape and statistical significance of the spectral lines are commented. Several criteria to determine the order of the prediction filter are considered.

Results include cases of white noise, autoregressive processes, sun-spot numbers and annual rainfall in the semiarid part of Chile. In cases accepting an autoregressive model the spectra depend weakly on the order of the filter, but strongly in cases of white noise and rain

fall data. In the later the criteria for order do not pin point any value, although some spectral lines appear to be significant, particularly lines at frequencies close to the quasibienial cycle.

As a general rule, it seems advisable to scan values for the filter order, since the significance is greatest at an intermediate value, which is not detected by the order criteria.

The sun-spot number series can be modelled by a third order autoregressive process.

1. INTRODUCCION.-

En el análisis de series de tiempo, la estimación del espectro de potencia es una práctica muy socorrida cuando se trata de identificar causas, relacionar dos fenómenos o simplemente describir el comportamiento de la serie.

La primera técnica que permitió su estimación apareció en los trabajos de N. Wiener publicados en los años de postguerra. Consiste en el cálculo de la Transformada de Fourier del autocorrelograma originado en la serie de observaciones. Para el cálculo de la transformada es conveniente introducir una función ponderante que anule en forma suave su valor para desfases elevados.

El descubrimiento de un algoritmo rápido para evaluar la Transformada de Fourier permitió aplicarla repetida y directamente a partes de la serie original para luego suavizar los resultados combinándolos. Aunque esta técnica no recurre a la función ponderante asume una serie infinita de observaciones merced a la repetición periódica de los datos originales.

Las dos técnicas mencionadas requieren suposiciones arbitrarias en relación a la continuación de la serie observada. El método de Wiener asume que la serie continúa idénticamente nula, pues sólo así es posible garantizar la nulidad de la función de autocorrelación, y el método que recurre a la transformada rápida, o método del periodograma, supone su repetición indefinida. Ambas son igualmente arbitrarias y afectan al espectro resultante.

Posteriormente, se ha perfeccionado una técnica que no requiere de suposición en cuanto al comportamiento de la serie más allá de lo observado. Esta se conoce como el Método de Máxima Entropía.

Las ventajas del Método de Máxima Entropía (MME) han sido verificadas en diversas áreas de la geofísica: geomagnetismo, movimientos del eje terrestre, meteorología, etc. Su poder resolutivo, superior a los métodos tradicionales, ha sido expuesto por Lacoss (1971) y Ulrich y Clayton (1976) para series sintéticas. En particular, cuando las series por analizar son cortas sus ventajas se acentúan.

Las bases teóricas del método han sido expuestas por Smylie et al. (1973), Andersen (1974), Burg (1975), Ulrich y Bishop (1975) y Johnsen y Andersen (1978), configurando un cuadro bastante completo, salvo la determinación del orden del filtro predictor necesario en el cálculo.

En estos trabajos se ha recurrido sistemáticamente a series sintéticas que en la mayoría de los casos son originadas en modelos autorregresivos o procesos que aceptan tal representación. El éxito del Método de Máxima Entropía no se extiende necesariamente a series de tiempo modelables por otros procesos. En particular, series modelables a través de esquemas más complejos, como procesos autorregresivos de promedio móvil o procesos no lineales, pueden no ser eficientemente analizados por esta técnica. Treitel et al. (1977) presentan ejemplos de series sintéticas lineales analizadas con diversas técnicas de estimación espectral concluyendo que el origen de la serie debe ser considerado en la elección del método de análisis espectral a usar. Para una serie de promedio móvil se recomienda el método de Wiener, para otra autorregresiva es apropiado el Método de Máxima Entropía y para series mixtas se debe usar una técnica iterativa que ellos presentan.

En este artículo se intenta mostrar el comportamiento del Método de Máxima Entropía en casos de series sintéticas y naturales, poniendo el acento en las dificultades que conlleva su utilización en ciertos casos donde los resultados son ambiguos. Se presentan algunos ejemplos que sirven para apreciar la influencia del nivel de ruido que cubre la señal y del orden o suavizamiento del método.

2. PRINCIPIOS BASICOS.-

Sean x_n , $n = 1, 2, \dots, (s-1)$ los valores conocidos de una serie para la cual se desea pronosticar el valor siguiente x_s en base a los N valores anteriores. Una manera de resolver el problema es a través del

diseño de un filtro lineal con coeficiente a_n que no proporcione una estimación \hat{x}_s de x_s a través de la relación

$$\hat{x}_s = - \sum_{n=1}^N a_n x_{s-n}$$

El error cometido será:

$$\epsilon_s = x_s - \hat{x}_s = x_s + \sum_{n=1}^N a_n x_{s-n}$$

$$\epsilon_s = \sum_{n=0}^N a_n x_{s-n} \quad \text{con } a_0 \equiv 1$$

En general, los errores no serán nulos y una aspiración natural es que la serie de errores carezca de información o estructura. Esto equivale a pretender que la serie de errores corresponda a un ruido blanco.

La última expresión muestra al filtro como un sistema que alimentado por la serie x_n , entrega la serie error de pronóstico que deseamos tenga un espectro blanco. A esta expresión del dominio temporal le corresponde otra en el dominio de la frecuencia, a saber

$$\zeta(z) = \left(\sum_{n=0}^N a_n z^{-n} \right) X(z)$$

donde $z = e^{j2\pi f \Delta t}$, f representa la frecuencia, Δt el intervalo de tiempo que separa dos valores sucesivos de la serie original, j es la unidad imaginaria, $\zeta(z)$ es la transformada zeta de la serie error ϵ_n y $X(z)$ la de la serie x_n .

Puesto que, salvo un factor constante, el producto $\zeta(z) X(z^{-1})$ es igual a la transformada zeta de la función de autocorrelación $R(n)$,

$$\zeta(z) \zeta^*(z) = \left(\sum_{n=0}^N a_n z^{-n} \right) \left(\sum_{m=0}^N a_m^* z^m \right) X(z) X^*(z)$$

$$P_\epsilon(f) = \sum_{n=0}^N a_n z^{-n} \sum_{m=0}^N a_m^* z^m P_X(f)$$

donde la substitución de la variable independiente z por la frecuencia f implica su evaluación sobre el círculo unitario $z = e^{j2\pi f \Delta t}$, $P_\epsilon(f)$ es la

densidad espectral de la serie error y $P_x(f)$ la correspondiente a la serie original. Luego:

$$P_x(f) = \frac{P_\epsilon(f)}{\sum_{m=0}^N a_m z^{-m} \sum_{n=0}^N a_n^* z^n}$$

Si el filtro a_n está tan bien diseñado que la serie error corresponde a ruido blanco con media nula y varianza σ^2 , su espectro vale $\sigma^2 \Delta t$ y resulta

$$P_x(f) = \frac{\sigma^2 \Delta t}{\sum_{m=0}^N a_m z^{-m} \sum_{n=0}^N a_n^* z^n} \quad (1)$$

Para obtener ecuaciones que permitan estimar los coeficientes a_n consideremos el error cuadrático del pronóstico:

$$\begin{aligned} \epsilon_s \epsilon_s^* &= \sum_{n=0}^N a_n x_{s-n} \sum_{r=0}^N a_r^* x_{s-r} & a_0 &\equiv 1. \\ &= \sum_{n=0}^N \sum_{r=0}^N a_n x_{s-n} x_{s-r} a_r^* \end{aligned}$$

El promedio de los errores cuadráticos de la serie de largo M será:

$$\begin{aligned} \frac{1}{N} \sum_{s=1}^{M-N} \epsilon_s \epsilon_s^* &= \sum_{n=0}^N \sum_{r=0}^N a_n a_r^* \frac{1}{N} \sum_{s=1}^{M-N} x_{s-n} x_{s-r} \\ &= \sum_{n=0}^N \sum_{r=0}^N a_n a_r^* R(n-r) \end{aligned}$$

donde $R(n)$ es la covarianza de la serie para el desfase n . Su minimización con respecto a los coeficientes a_n , $n = 1, 2, \dots, N$ conduce a N ecuaciones de la forma

$$\sum_{s=0}^N R(n-s) a_s = 0 \quad n = 1, 2, \dots, N$$

pero no es posible minimizar en relación a a_0 que está determinado ($a_0 \equiv 1$). Una relación análoga a las anteriores no se anula para $n=0$ y se tendrá:

$$\sum_{s=0}^N R(-s)a_s = \sum_{s=0}^N R(s)a_s = P_N \quad (2)$$

El sistema de $N+1$ ecuaciones

$$\sum_{s=0}^N R(n-s)a_s = P_N \delta_{0n} \quad n = 0, 1, 2, \dots, N \quad (3)$$

donde δ_{ij} es la función delta de Kronecker, es suficiente para proveer los valores de P_N y a_1, a_2, \dots, a_N .

La cantidad P_N es real y positiva como consecuencia del carácter Hermitiano de la matriz de autocovarianza $R(n-s)$ de la ecuación (3), pues multiplicando esta relación por el vector $a_r^* = (1, a_1^*, a_2^*, \dots, a_N^*)$ resulta

$$\sum_{r=0}^N \sum_{s=0}^N a_s R(n-s) a_r^* = P_N$$

y puesto que la matriz de autocovarianza es positiva definida $P_N > 0$ y real.

Una interpretación de P_N se puede obtener de la siguiente manera. Sean $1, a_1, a_2, \dots, a_N$ la solución del sistema (3), $1, a_1', a_2', \dots, a_N'$ las componentes de un filtro de pronóstico arbitrario y consideremos el valor H , positivo o nulo que corresponde a la expresión

$$\sum_{m=0}^N \sum_{n=0}^N (a_m'^* - a_m) R(m-n) (a_n' - a_n) \equiv H$$

Haciendo uso de (3) se reduce a:

$$H = \sum_{m=0}^N \sum_{n=0}^N a_m'^* R(m-n) a_n' - P_N$$

Cuando $a_n' = a_n$, $n = 1, 2, \dots, N$, $H = 0$ y la doble sumatoria, que es el error cuadrático medio del pronóstico, se hace mínima e igual a P_N .

Si la serie original proviene de un proceso autorregresivo de la forma

$$x_r = \sum_{n=1}^N a_n x_{r-n} + z_r$$

donde z_r es una componente aleatoria con valor medio nulo y varianza σ^2 , el error de pronóstico coincide con z_r y la correlación

$$\frac{1}{N} \sum_{r=1}^N x_r \varepsilon_r = \frac{1}{N} \sum_{r=1}^N z_r^2 = \frac{1}{N} \sum_{r=1}^N \varepsilon_r^2 = \sigma^2 = P_N$$

se identifica con la varianza del error de pronóstico mínimo.

Sin embargo, en general la serie de tiempo puede no corresponder a un proceso autorregresivo o bien el orden N usado no coincidir con el correcto, de modo que en el estimador de la densidad espectral conviene substituir su valor por P_N , y la relación (1) da origen a:

$$P_x(f) = \frac{P_N \Delta t}{\sum_{m=0}^N a_m z^{-m} \sum_{n=0}^N a_n^* z^n} \quad (5)$$

estimador espectral que se conoce como de Máxima Entropía.

De lo expuesto se infiere que el Método de Máxima Entropía entregará un estimador espectral óptimo cuando la serie corresponda a un proceso autorregresivo y se haya averiguado su orden. Más adelante veremos que existen algunos criterios para tal efecto, pero desde ya es posible deducir que el comportamiento de P_N como función del orden N será uno de ellos. Resulta lógico que a medida que se agregan términos al filtro la correlación error-serie disminuya si ellos dan cuenta de información contenida en la serie, pero cuando se exceda el orden correcto la disminución ya no será significativa.

Entonces el análisis espectral por el Método de Máxima Entropía puede ser considerado como una técnica que determina el proceso autorregresivo que se ajusta mejor a los datos y luego evalúa su espectro. Desde este punto de vista la expresión (5) es una relación analítica que puede evaluarse con una densidad de puntos arbitrariamente grande, lo que permite una mejor definición de las líneas espectrales. Resulta natural así, que

el Método de Máxima Entropía ofrezca una excelente resolución cuando el ajuste es factible. Por otra parte el ajuste puede ser considerado un proceso de filtrado, análogo al uso de ventanas y promedios en las técnicas convencionales, pero en este caso el filtro está determinado por los datos y no obedece a una elección arbitraria.

La solución del sistema (3) se facilita por la simetría de su matriz. Levinson (1947) ideó un algoritmo que itera sobre el orden del sistema, que ha sido revisado recientemente por Burg (1968). En la versión original el algoritmo precisa de la evaluación previa de la función de autocovarianza, $R(\tau)$, para desfases menores e iguales al orden N . Esto se debe realizar en base a la serie de datos que se analiza. Para series cortas los estimadores corrientes

$$R_1(\tau) = \frac{1}{M} \sum_{n=0}^{M-\tau} x_n^* x_{n+\tau} \quad ; \quad R_2(\tau) = \sum_{n=0}^{M-\tau} x_n^* x_{n+\tau}$$

son insatisfactorios pues el primero es sesgado y el segundo puede dar lugar a una matriz de autocovarianza que no sea positiva definida. La innovación de Burg evita la estimación de $R(\tau)$ y con ella sus deficiencias para casos de series cortas. La técnica está descrita con detalle en el trabajo de Ulrich y Bishop (1975), donde se hace ver que los estimadores \hat{R}_1 y \hat{R}_2 suponen que x_n es nulo para $|n| > M$ en tanto que en la técnica de Burg esta extensión arbitraria de los datos está ausente, pues en ella el error cuadrático del filtro se obtiene haciendo su convolución con la serie de datos ordenados en forma directa e inversa. La función de autocovarianza que se obtiene del proceso no compromete valores fuera de los datos y conforma una matriz positiva definida.

Las experiencias que se describen en la tercera parte del trabajo sugieren que el Método de Máxima Entropía tiene algunos inconvenientes serios. El más importante es la carencia de un criterio objetivo para estimar el orden del filtro. Cuando la serie es representable por un proceso autorregresivo (PAR) existen criterios que se describen en la Sección 2.2, pero cuando este no es el caso fue preciso barrer órdenes y observar el comportamiento del espectro. Buscando algún indicio definitorio se realizaron experiencias con series sintéticas con los resultados que se describen más adelante. En esta búsqueda se apreció la falta de un estimador de la varianza del espectro que permita separar rasgos significativos, aun que existe una solución aproximada sugerida por Tabony (1979) no parece ser del todo decisiva. Se la describe brevemente en la Sección 2.2.

2.1 Potencia y ancho de líneas espectrales.

El análisis de la expresión (1) en el plano complejo es particularmente esclarecedor. Notemos, en primer lugar, que si $z = e^{j2\pi f\Delta t}$ su lugar geométrico es el círculo unitario. No obstante, es posible considerar la continuación analítica de la función representada en (1) desde el círculo a todo el plano complejo. Puesto que los coeficientes a_n son reales, el denominador de la relación presentará raíces reales o complejas conjugadas. Una pareja de raíces que se ubiquen sobre el círculo unitario implica la presencia de una componente armónica por el ángulo polar de la raíz ϕ_0 , pues $\phi_0 = \pm 2\pi f_0 \Delta t$. Este resultado es consecuencia inmediata del método de solución de ecuaciones de diferencias a través de las raíces del polinomio característico (ver por ejemplo Henrici, 1964). En la densidad espectral esto se reflejará en un máximo muy agudo, en estricto rigor la función $P(f)$ crece sin medida a la frecuencia correspondiente a f_0 .

El denominador de la expresión (1) puede transformarse de la siguiente forma:

$$\sum_{m=0}^N a_m z^{-m} \sum_{n=0}^N a_n z^n = \sum_{m=0}^N a_m z^{-m+N} \sum_{n=0}^N a_n z^{n-N}$$

La ecuación $A(z) \equiv \sum_{m=0}^N a_m z^{N-m} = 0$ es la ecuación característica del filtro predictivo. Su estabilidad se debe reflejar en que todas sus raíces, $z_k = r_k e^{j2\pi f_k \Delta t}$, se ubiquen dentro del círculo unitario, $|r_k| < 1$. El segundo factor $A(\frac{1}{z}) = \sum_{n=0}^N a_n z^{-(N-n)}$ tendrá sus raíces, $\frac{1}{z_k^*} = \frac{1}{r_k} e^{+j2\pi f_k \Delta t}$ fuera del círculo unitario.

La expresión (5) del espectro, muestra que para un orden N fijo, los máximos de la función $P_x(f)$ se originan cuando una raíz de $A(z)$ se encuentra próxima al círculo unitario. Johnsen y Andersen (1978) demostraron que el área bajo la curva espectral se podía expresar por

$$\int_{-1/2}^{1/2} P_x(f) df = \frac{1}{2\pi j} \oint \frac{P_x(z)}{z} dz$$

merced a la relación $z = e^{j2\pi f\Delta t}$, y que por el Teorema del Residuo será igual a la suma de los residuos de $P_x(z)/z$ evaluados en $z = z_k$, las raíces de la ecuación característica. Por este procedimiento es posible asignar a cada línea espectral una potencia, o fracción de la varianza de la serie, igual al residuo correspondiente del polo que genera la línea. Adicionalmente los mismos autores demostraron que las líneas del espectro de máxima entropía tienen una forma de Lorentz con ancho de banda, w_k , igual a $(\frac{1}{r_k} - r_k)/2\pi$, esto es

$$P_x(f) = \frac{P_x(f)}{1 + \left(2 \frac{f - f_k}{w_k}\right)^2}$$

donde f_k es la frecuencia correspondiente al máximo de la línea.

2.2 Criterios para determinar el orden.-

No deja de llamar la atención el que J.P. Burg no mencione en sus trabajos el problema que impone la elección del orden del modelo a ajustar. Sin embargo, se trata de un aspecto importantísimo pues determina la suavidad de la distribución espectral. Ulrich y Bishop (1975) sugieren que dos criterios propuestos por Akaike (1970 y 1974) podrían ser útiles: el primero es el error final de pronóstico (FPE) y el segundo se conoce como el criterio teórico de información (AIC).

Ya se ha comentado que un buen criterio debe ser el valor del error cuadrático medio mínimo P_N , cuyo valor decrece rápidamente cuando la adición de un nuevo término al modelo representa una incorporación importante de información. Los criterios de Akaike dependen del valor de P_N . Sin entrar en su justificación las expresiones son

$$(FPE)_N = \frac{M + (N + 1)}{M - (N + 1)} P_N$$

$$(AIC)_N = M \log P_N + 2(N + 1)$$

donde M es el número de términos en la serie de observaciones y N el orden del filtro. En ambas expresiones se advierte la combinación de dos efectos opuestos; por un lado P_N decrece con el aumento del orden N y por otro el factor y el sumando del FPE y AIC respectivamente, crecen con él, de modo que ambas expresiones deben pasar por un valor mínimo que se considera que identifica el orden apropiado. Uno de los revisores nos ha hecho notar que el criterio AIC ha sido extendido exitosamente para series que aceptan representaciones mixtas (autorregresivas con promedio móvil), por el mismo Akaike (1979).

2.3 Significancia de líneas espectrales.-

Tabony (1979) ha presentado una técnica aproximada para estimar la significancia de los máximos espectrales. En ella se limitan las líneas por las frecuencias que se ubican a niveles de $1/3$ del valor máximo, se calcula la fracción de varianza incluida entre ellas y se compara con la fracción correspondiente a un espectro blanco. Suponiendo que no existe correlación entre puntos vecinos se asociaron a la varianza total un número de grados de libertad igual al número de observaciones. Asumiendo que los grados de libertad se distribuyen parejamente entre las distintas frecuencias se asocia a la línea una fracción correspondiente a su ancho entre los puntos de $1/3$ de potencia. Con ello se aplica un test de Fisher a la razón de varianzas.

Para una línea con perfil de Lorentz los puntos de $1/3$ de potencia máxima están separados por $2W_k$, siendo W_k el ancho de banda entre las frecuencias de $1/2$ de potencia máxima. La potencia entre los puntos $1/3$ de potencia máxima es $0.705 P$ siendo P la potencia total de la línea. El procedimiento de Tabony fue aplicado a la línea de Lorentz descrita en la Sección 2.1.

3. APLICACIONES DEL METODO DE MAXIMA ENTROPIA.-

Para montar un programa que permita usar los resultados anteriores se usó una subrutina de Ulrich y Bishop (1975) como base. A ellas se agregaron subprogramas que calculan las raíces de la ecuación característica y evalúan la potencia y ancho de la banda de cada línea espectral.

Este programa fue aplicado a series sintéticas y series de observaciones. Entre las sintéticas se usaron series de ruido gaussiano puro y series generadas por procesos autorregresivos. Entre las secuencias de observaciones se usó una serie de números de Wolfer, relacionados con las manchas solares, cuya periodicidad es bien conocida. Adicionalmente se analizaron series pluviométricas anuales de estaciones chilenas que se caracterizan por una gran cantidad de ruido y ausencia de periodicidades conocidas.

Con estas elecciones se intenta extraer una impresión de las bondades y defectos del método a través de la comparación de resultados para series sintéticas y reales.

3.1 Series de ruido blanco.-

La generación sintética de series de ruido blanco gaussiano se realizó por medio de una subrutina que genera números aleatorios. En todos los casos se usó un valor medio cero, pero diferentes desviaciones standard a fin de generar distintas intensidades de ruido.

La Fig. 1 presenta gráficos correspondientes a una serie de ruido blanco gaussiano con varianza igual a cuatro. En ella se incluye una muestra de la serie original, x_r , en función de r ; parte de la función de autocorrelación obtenida por el método de Burg, $R(n)$, en función del desfase n ; el criterio AIC de Akaike, el error cuadrático medio mínimo P_N y el coeficiente de correlación parcial a_N , todos en función del orden N del modelo. La Fig. 2 muestra los espectros de máxima entropía para modelos de orden 1.9 y 32.

Los resultados muestran que aunque se trata de ruido blanco el espectro de potencia exhibe un número de líneas que depende del orden del filtro ajustado. Además, a medida que aumenta el número de máximos espectrales estos se hacen más agudos y elevados dificultando el reconocimiento de un espectro blanco. Lo que mejor caracteriza la ausencia de estructura en la serie son las curvas correspondientes de órdenes bajos, uno y dos. En particular el filtro de primer orden muestra un valor casi constante. Esto implica que la única raíz de su ecuación característica

$$1 + a_1 z^{-1} = 0$$

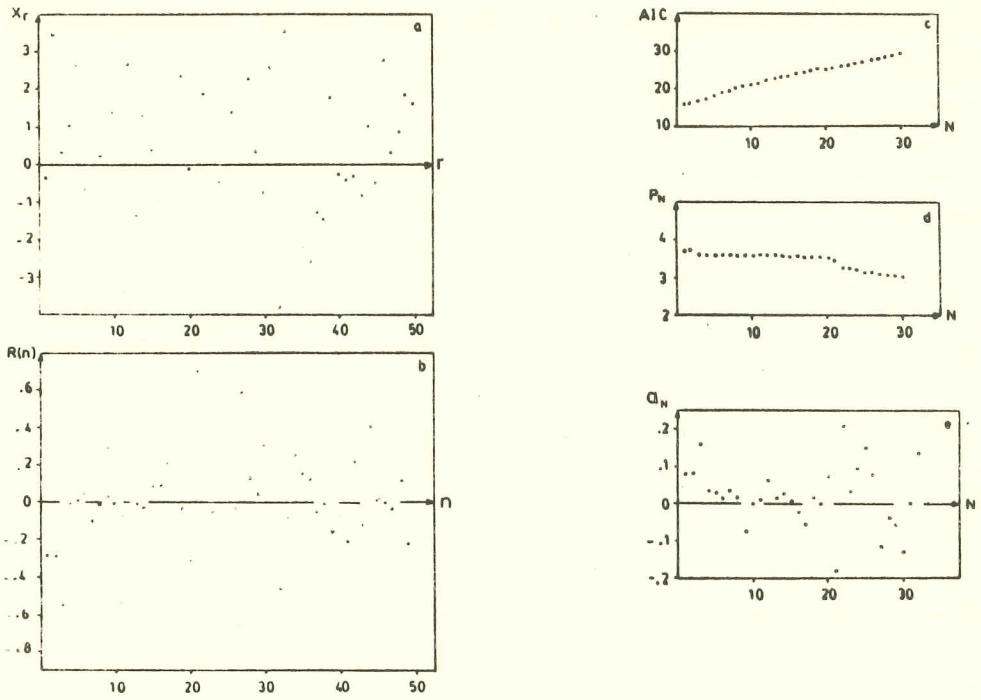


FIG. 1: Ruido gaussiano puro. (a) Gráfico de la serie original. (b) Función de autocorrelación en función del desfase. (c) AIC de Akaike en función del orden. (d) Error cuadrático medio mínimo del filtro predictor en función del orden del filtro. (e) Coeficiente de correlación parcial en función del orden del modelo ajustado.

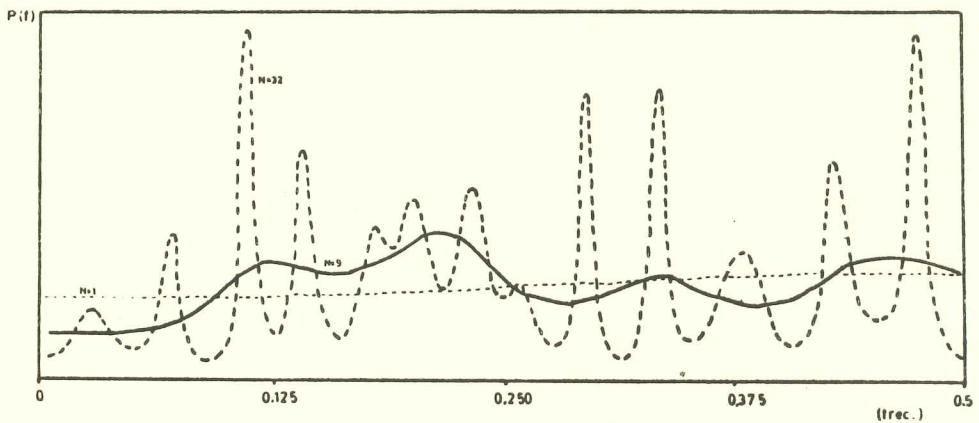


FIG. 2: Densidad espectral estimada por el método de Máxima Entropía para una serie de ruido gaussiano puro con modelos ajustados de órdenes 1, 9 y 32.

se ubica próximo al origen en el plano complejo z , consecuencia directa del pequeño valor de la autocorrelación para un desfase unitario.

Otras realizaciones de ruido blanco con igual varianza entregan estructuras y líneas diferentes manteniendo un valor aproximadamente invariables en el caso del filtro de orden uno.

Como es lógico, un aumento en la varianza del ruido sube el nivel medio del espectro, pero las características comentadas permanecen inalteradas.

Ninguna de las líneas evidenciadas resulta significativa al nivel 10%. La línea que aporta más a la varianza con $N = 9$, es decir cinco líneas posibles, concurre con un 28% y corresponde a la frecuencia 0.218.

Los gráficos del criterio AIC y de P_N , en la Fig. 1, son muy parejos y sólo muestran un pequeño quiebre para el orden $N = 20$.

3.2 Procesos autorregresivos.

La síntesis de procesos autorregresivos de orden N definidos por la relación

$$a_r = a_1 x_{r-1} + a_2 x_{r-2} + \dots + a_N x_{r-N} + \epsilon_r$$

se llevó a cabo usando la subrutina de números aleatorios ya mencionada para proveer los elementos aleatorios ϵ_r . El proceso se inicializó con valores nulos y se dejó correr 200 pasos para alcanzar un estado de régimen. En seguida se generaron 100 valores que se sometieron a análisis. Se trabajó con procesos autorregresivos de orden dos y cuatro.

Se usaron dos procesos de segundo orden, el primero tomado de Ulrich y Bishop (1975) sirvió para verificar los programas y obedece a la relación

$$x_r = 0.75 x_{r-1} - 0.5 x_{r-2} + \epsilon_r \quad (4)$$

teniendo ϵ_r una distribución gaussiana con valor esperado nulo y varianza unitaria. Las raíces de la ecuación característica son $z_{1,2} = (3 \pm j\sqrt{23})/8$

cuyos módulos son menores que uno. La frecuencia correspondiente al máximo resulta ser $f_{12} = 0.161$.

La serie y su función de autocorrelación se muestran en la Fig. 3. Su espectro para órdenes 4, 9 y 49 se representan en la Fig. 4.

En este caso resulta un espectro extraordinariamente estable, los órdenes 2, 3, 4 y 9 dan lugar a curvas muy parecidas. Sólo difieren en la magnitud del máximo y levemente en su frecuencia. Cuando el orden aumenta a 49 aparecen variaciones espúreas que oscilan en torno al espectro real sin llegar a encubrirlo. Por lo tanto, para este caso existe un intervalo de órdenes que proveen la información buscada y sólo cuando se la excede apreciablemente se oscurece la interpretación.

Las Figs. 5 y 6 ilustran el mismo proceso de las Figs. 3 y 4 con una varianza mayor ($\sigma^2 = 100$). Nótese como el filtro de primer orden origina un máximo en las frecuencias bajas, contrastando con los casos de ruido blanco donde se generaba un valor casi constante. Los gráficos del AIC y de P_N ponen en evidencia la bondad de ambos en la determinación del orden. El AIC pasa por un mínimo para $N = 2$ y P_N cae bruscamente entre $N=1$ y $N=2$, no hay duda que el proceso que originó la serie es de segundo orden. Los coeficientes a_N también logran valores importantes para $N=1$ y 2 para oscilar con pequeña amplitud cuando $N > 2$.

También se recurrió al proceso autorregresivo de cuarto orden.

$$x_r = 2.7607 x_{r-1} - 3.8106 x_{r-2} + 2.6535 x_{r-3} - 0.9238 x_{r-4} + \epsilon_r$$

que ha sido usado en la literatura para mostrar el poder resolutivo de distintos métodos, pues tiene dos máximos muy próximos. Se usaron órdenes 1, 2, 9 y 49 todos los cuales entregaron un espectro muy estable debido a que los ceros de la ecuación característica se aproximan al círculo unitario y las perturbaciones de ceros espúreas tienen magnitud relativa muy pequeña.

Finalmente se consideró un caso de segundo orden en que los ceros de la ecuación característica estaban sobre el círculo unitario, o sea un proceso armónico distorsionado con diferentes niveles de ruido (varianza 1, 16 y 100). El proceso

$$x_r = 1.732 x_{r-1} - x_{r-2} + \epsilon_r$$

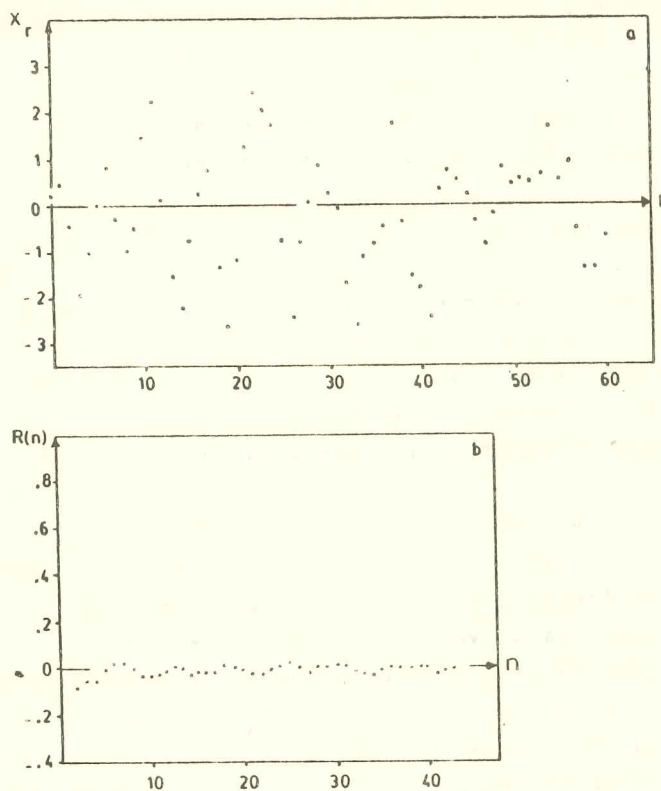


FIG. 3: Proceso autorregresivo de segundo orden. (a) Gráfico de la función original. (b) Función de autocorrelación en función del desfase.

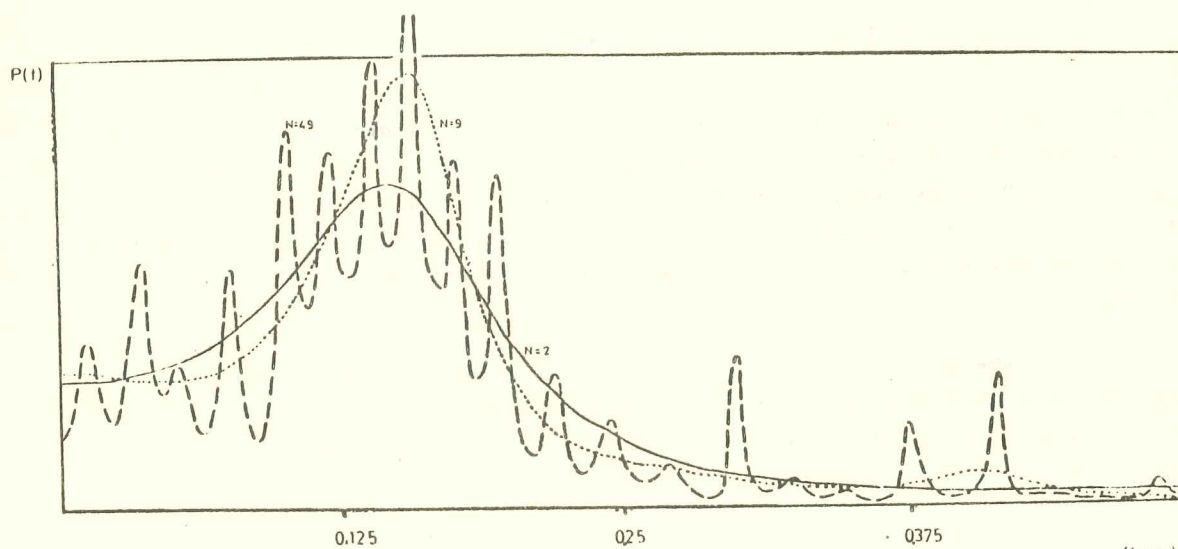


FIG. 4: Proceso autorregresivo de segundo orden: densidad espectral estimada por el método de Máxima Entropía con modelos de órdenes 1, 9 y 49.

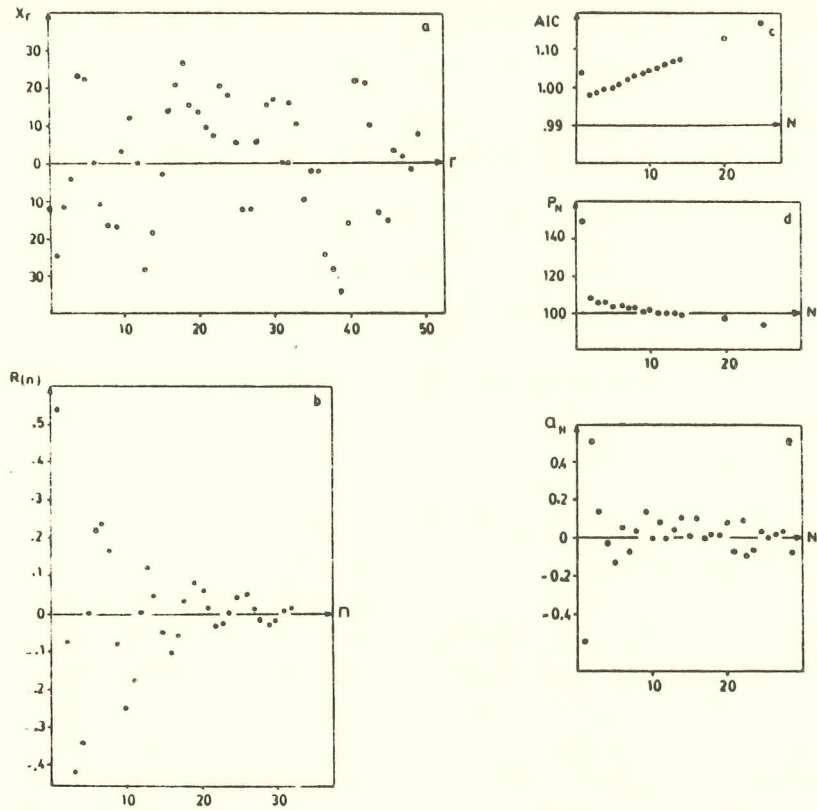


FIG. 5: Proceso armónico con ruido gaussiano. (a) Gráfico de la función original. (b) Función de autocorrelación en función del desfase. (c) AIC de Akaike en función del orden. (d) Error cuadrático medio mínimo en función del orden del filtro. (e) Coeficiente de correlación parcial en función del orden del modelo ajustado.

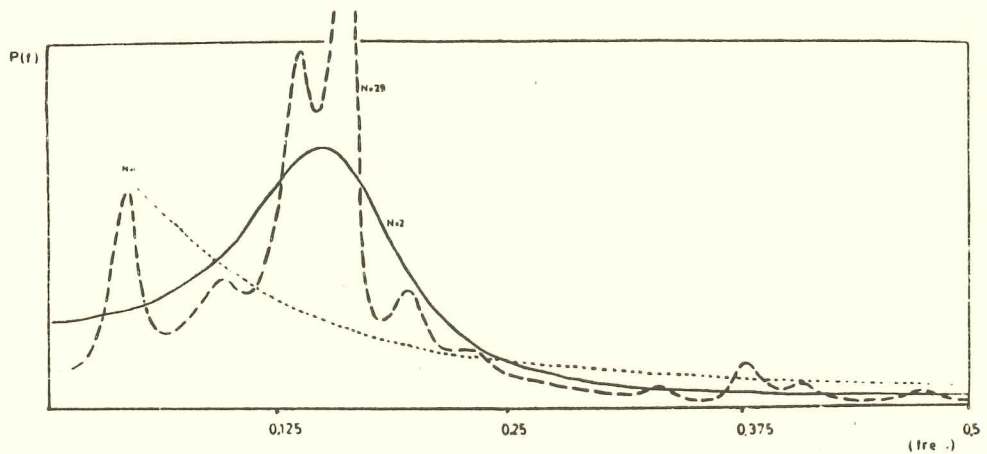


FIG. 6: Proceso armónico con ruido gaussiano: densidad espectral estimada por el Método de Máxima Entropía con modelos de órdenes 1, 2 y 29.

tiene por raíces $z_{1,2} = 0.886 \pm 0.5 j$, y su espectro entrega un máximo muy agudo en la frecuencia 0.0833. Una mayor varianza del ruido sólo hace aumentar la escala, pero los máximos secundarios no llegan a perturbar la suavidad del espectro calculado.

En todos estos casos los máximos de órdenes intermedios y bajos resultan significativos a un nivel de 5%.

3.3 Serie de manchas solares.

Entrando ya en el análisis de series geofísicas se recurrió primero a una de 110 términos que representa el número de manchas solares entre los años 1867 y 1976. Sus valores se obtuvieron de Menzel (1959) y los últimos años fueron proporcionados por el Departamento de Astronomía de esta Facultad. La serie de manchas solares es probablemente una de las más analizadas. Es bien conocido su ciclo de 11,2 años, cifra que corresponde a un valor medio de períodos que fluctúan entre 7.5 y 16 años.

La Fig. 7 incluye un gráfico de la serie original entre 1867 y 1920, (donde se han usado las desviaciones respecto de la media del período completo) la función de autocorrelación, el error cuadrático medio mínimo y las correlaciones parciales. La periodicidad de la serie es evidente en la serie misma y en su función de autocorrelación. Por otra parte las variaciones de P_N y a_N , sugieren que el proceso puede ser descrito óptimamente con un modelo de tercer orden, pues luego de $N=3$ el error se estabiliza y los coeficientes a_N disminuyen a valores pequeños.

El proceso ajustado está regido por la relación

$$x_r = 1.19625 x_{r-1} - 0.39246 x_{r-2} + 0.17949 x_{r-3} + \epsilon_r$$

donde ϵ_r es una variable aleatoria normal con media cero y varianza $\sigma_\epsilon^2 = 371$.

El espectro de máxima entropía de la serie, Fig. 8, evidencia un máximo único centrado en un período de 11,1 años. El aumento del orden de 3 a 14 sólo mejora la magnitud del máximo, pero un orden de 29 ya contiene oscilaciones ficticias. En la misma figura se puede apreciar el comportamiento del espectro de orden uno, el que se concentra en las frecuencias bajas.

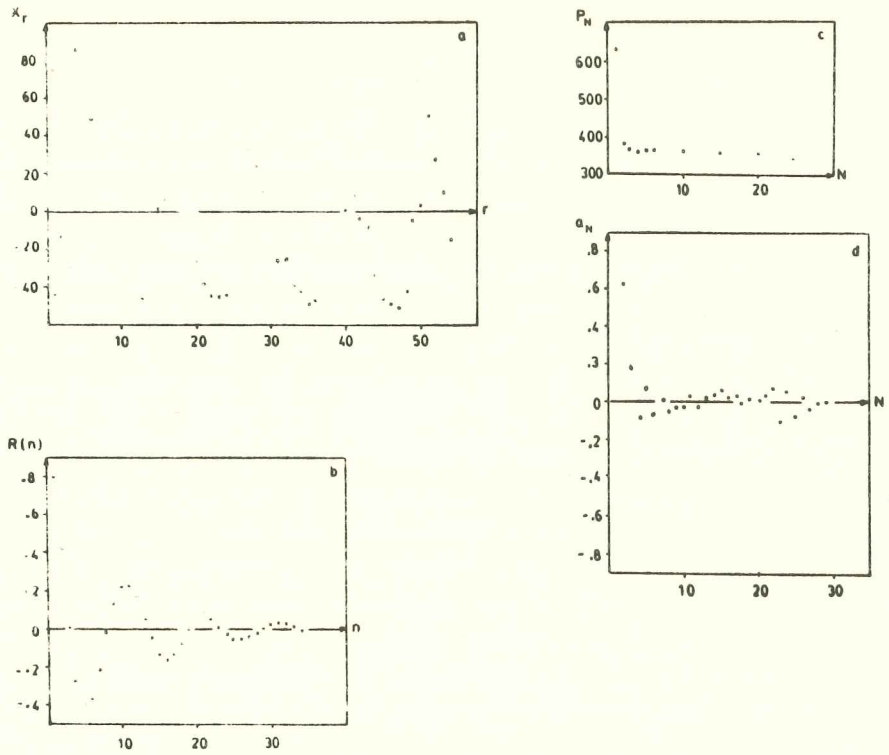


FIG. 7: Serie de números de manchas solares. (a) Gráfico de las desviaciones con respecto del valor medio (54 valores iniciales). (b) Función de autocorrelación en función del desfase. (c) Error cuadrático medio mínimo en función del orden. (d) Coeficiente de correlación parcial en función del desfase.

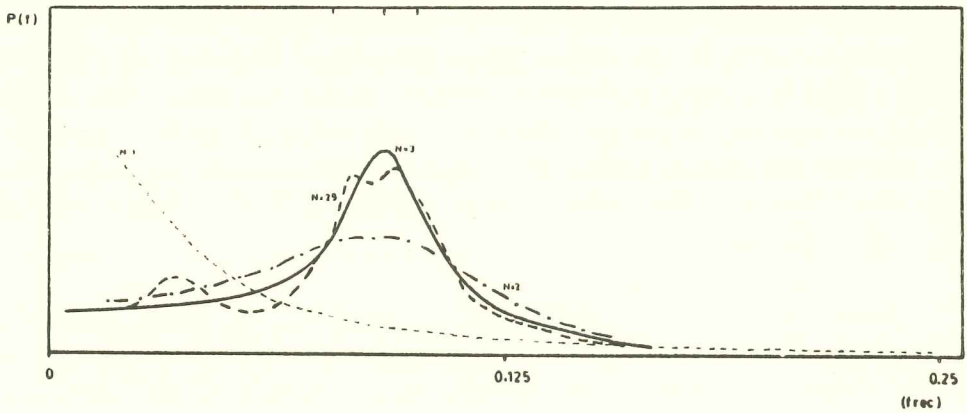


FIG. 8: Serie de números de manchas solares: densidad espectral estimada por el método de Máxima Entropía con modelos de órdenes 1, 2, 3 y 29. Las marcas en el borde superior identifican períodos de 13.3, 11.1 y 10 años.

Cuando a la serie original de manchas solares se le agrega ruido blanco gaussiano la serie resultante va perdiendo sus características periódicas y para cierto nivel de ruido (σ^2 aproximadamente 100) ya es imposible develar su comportamiento. Cuando el ruido agregado tiene una varianza igual a 50, el orden de ajuste óptimo se ha reducido desde el tercero al primero, el espectro de orden uno muestra un máximo muy vago en las frecuencias bajas, pero el máximo aun es bien evidenciado en órdenes altos (14 y 29) no así en el orden tres. Para ruido con varianzas igual a 100 las líneas espectrales ficticias son de magnitud similar a los del período de la serie sin ruido y los espectros de orden bajo son propios del ruido blanco.

Para $N=14$ y en ausencia de ruido artificial el único máximo significativo al 5% es el de 10.7 años. Al introducir ruido la significancia de la línea disminuye lentamente, a punto que con ruido de $\sigma^2 = 150$ tiene un valor marginal.

3.4 Serías pluviométricas de Chile Central.-

La búsqueda de periodicidades ocultas en series climatológicas tiene una larga historia (Mason, 1976) que en la mayoría de los ejemplos ha concluido en desengaños. La excepción más notable la constituye la oscilación cuasibienal de la estratósfera inferior tropical, donde el viento zonal medio y la temperatura oscilan con un período algo irregular, que alcanza a 26 meses en valor medio. La oscilación se propaga desde niveles superiores haciéndose imperceptible bajo 20 Km de altura. Sin embargo, existen evidencias parciales de su presencia cerca de la superficie (Wright, 1968 y Trenberth, 1975 y 1976). Entre los años 1950 y 1963 presentó un período de 26 meses, que luego aumentó a 30 meses, Holton (1975). También existen evidencias sujetas a controversia de períodos de 23 y 11.5 años, Mason (1976), todas ellas asociadas con pequeñas fracciones de la varianza total.

Entre las series climatológicas disponibles de Chile las más largas corresponden a las pluviométricas. A priori se consideró que las posibles periodicidades ocultas en las series de precipitación se manifestarían más claramente en la región semiárida subtropical. En esta zona La Serena posee el registro más largo. En consecuencia se sometió a análisis las sumas anuales acumuladas en esta ciudad desde 1870 a 1976. Las series analizadas corresponden al logaritmo de las sumas anuales expresadas en mm.

La Fig. 9 incluye parte de la serie, la función de autocorrelación, el error cuadrático medio del filtro P_N , el AIC de Akaike y las correlaciones parciales a_N . No es posible apreciar evidencia alguna de periodicidad o de preferencia por algún orden en particular. El espectro de máxima entropía para órdenes 1, 19 y 34 se muestra en la Fig. 10. El máximo mayor se ubica en el período de 2.2 años, coincidiendo con la oscilación cuasibienal, además muestra valores elevados en las frecuencias bajas. Posteriormente se analizaron series parciales correspondientes a los períodos 1870-1899, 1900-1929, 1930-1959 y 1940-1976, Fig. 11. El máximo en 2.2 años apareció en el segundo y cuarto período, estando completamente ausente en el primero y desplazado hacia el límite de 2 años en el tercero (todo con un orden 14). Por otra parte el espectro de baja frecuencia recibe un aporte fuerte en el período 1900-1929 cuando existió una fuerte caída de las precipitaciones, Aceituno (1977). En resumen, de la comparación de las Figs. 10 y 11 se aprecia que el espectro de la serie completa es una superposición de líneas de los períodos parciales, cada uno de los cuales muestra características distintas con la posible excepción del período cuasibienal.

Las pruebas de significancia a un nivel de 5% confirman el ciclo cuasibienal en La Serena cuando el orden es 9, 11 y 19. En este último caso también resulta marginalmente significativo un período de 7.3 años. Considerando sólo el ciclo de 2.2 años, su significancia es óptima para órdenes alrededor de 15. Cuando el orden del modelo autorregresivo es demasiado bajo la línea espectral es ancha aportando un exceso modesto sobre la varianza del ruido blanco. Cuando el orden es demasiado elevado la línea es muy aguda disminuyendo R y aumentando F por la pérdida de libertad.

A continuación se sometieron a análisis series pluviométricas de otras estaciones del Norte Chico: Caldera (1900-1967), Punta Tortuga (1900-1965) y Puerto Oscuro (1910-1976), Fig. 12. Todas ellas mostraron máximos en las vecindades de 2.2, 4.3 y 11 años, acompañadas de otras variables.

Con órdenes 9 y 19 sólo Punta Tortuga confirmó el período de 2.2 años, para el resto el máximo no es significativo al nivel 5%. Es posible que con un orden 15 lo sea.

Estaciones del centro y sur del país, como Juan Fernández, Valparaíso, Santiago, Concepción y Valdivia no mostraron evidencias de estos máximos espectrales, salvo Valparaíso que exhibe máximos en los períodos de 4.3 y 2.2 años que no resultaron significativos al 5%.

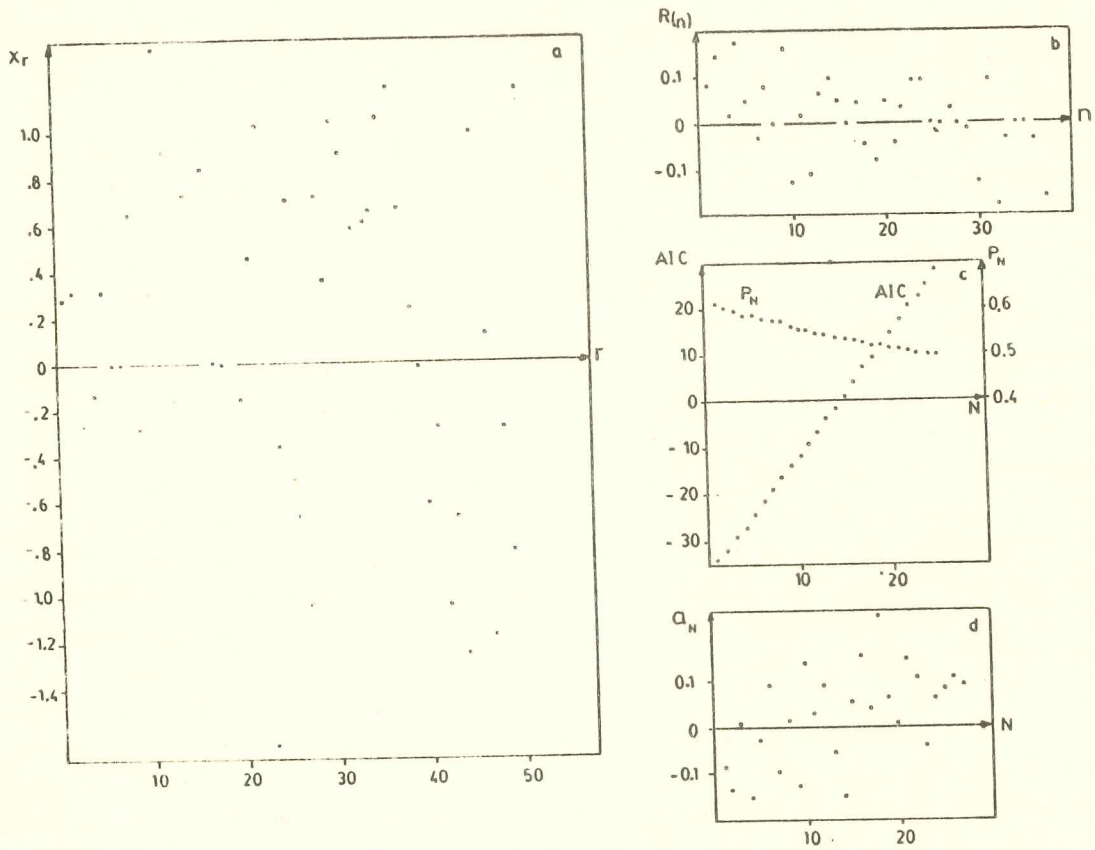


FIG. 9: Serie de pluviometría anual de La Serena. (a) Desviaciones con respecto al valor medio, del logaritmo de las precipitaciones anuales. (b) Función de autocorrelación en función del desfase. (c) Error cuadrático medio mínimo, P_N , y valor del AIC de Akaike en función del orden. (d) Coeficiente de correlación parcial en función del orden.

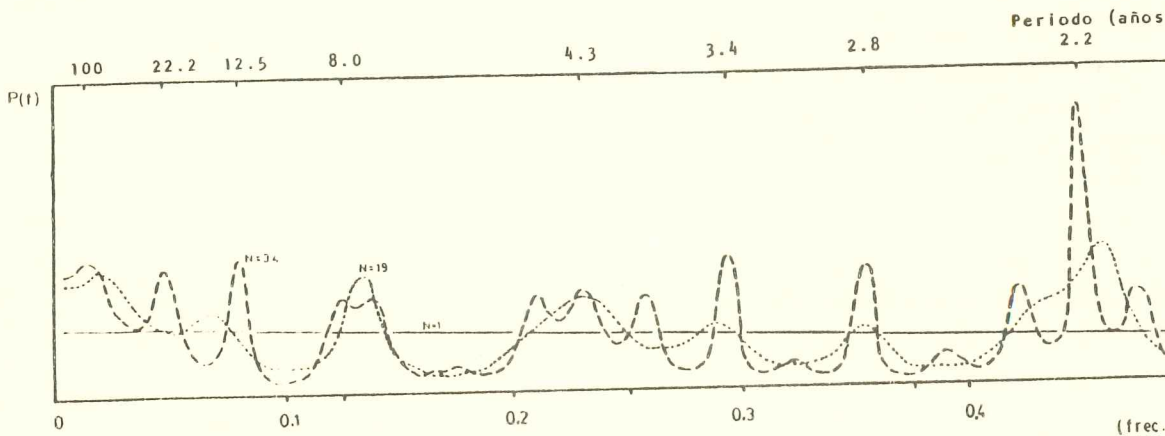


FIG. 10: Serie de pluviometría anual de La Serena: densidad espectral estimada por el Método de Máxima Entropía con modelos de órdenes 1, 19 y 34.

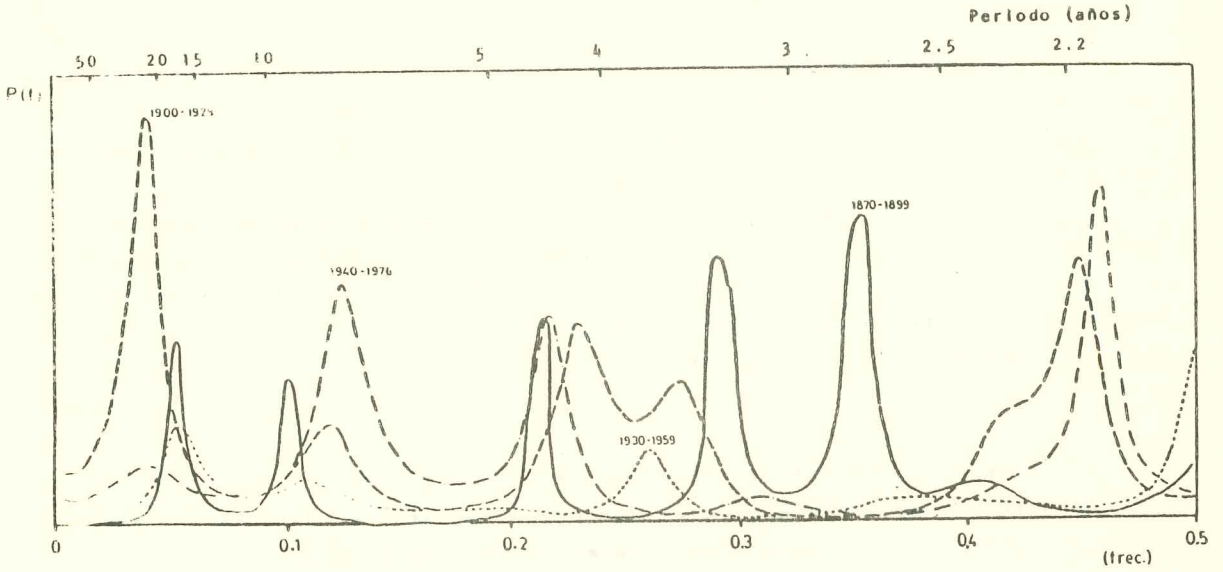


FIG. 11. Series pluviométricas parciales de La Serena: densidades espectrales estimadas por el Método de Máxima Entropía para diferentes períodos ajustando modelos de orden 14.

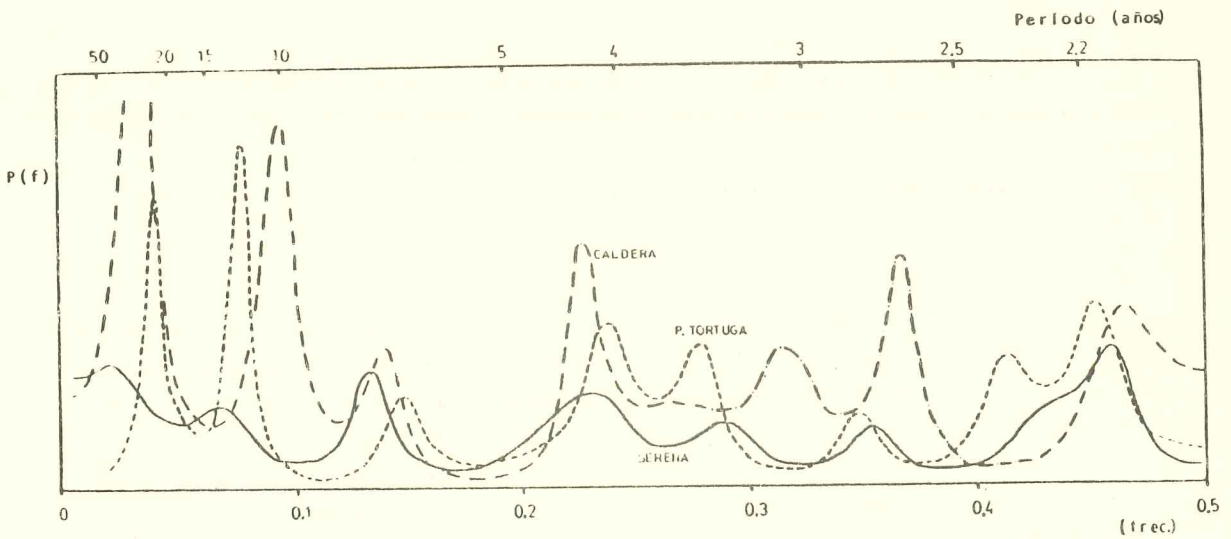


FIG. 12: Series pluviométricas del Norte Chico: densidades espectrales estimadas por el Método de Máxima Entropía.

4. CONCLUSIONES.-

El Método de Máxima Entropía provee estimaciones del espectro de potencia que dependen en un grado importante del orden del modelo o filtro ajustado. Solamente en el caso de series sintéticas autorregresivas entrega estimaciones estables, poco dependientes del orden usado y del nivel de ruido presente en la serie. Cuando se ignora el tipo de proceso que origina la serie, como sucede normalmente en la práctica, es aconsejable usar criterios de significancia para separar líneas reales de las espúreas. El criterio aproximado de Tabony (1979) parece ser adecuado según se desprende de los resultados obtenidos en las secciones 3.1, 3.2 y 3.3.

En casos de series con altos niveles de ruido la significancia de una determinada línea espectral parece optimizarse en órdenes intermedios (Sección 3.4) lo que obliga a hacer un barrido de órdenes.

Un producto derivado del análisis espectral por el Método de Máxima Entropía es la modelación de la serie por un proceso autorregresivo. En el caso de la serie de manchas solares el modelo obtenido está bien definido.

En las series pluviométricas de valores anuales de Chile Central los espectros resultantes son muy variables. Sólo en la región semiárida aparecen algunos máximos significativos entre los cuales destaca el período cuasibienal, que en La Serena, donde se hizo el estudio más detenido, es el mayor.

El estudio de series parciales, de La Serena, muestran la impermanencia de los rasgos espectrales en este tipo de series, de modo que en períodos del orden de 30 años la precipitación carece de estacionaridad. En cualquier caso, la proximidad del período de 2.2 años a la frecuencia de Nyquist (2 años) hace recomendable un intervalo de muestreo menor para evitar peligros de 'aliasing' y la falta de estacionaridad sugiere la extensión del período de muestreo. Finalmente, no se debe olvidar que es posible que el modelo implicado en el método no sea el más adecuado para el fenómeno.

5. REFERENCIAS BIBLIOGRAFICAS.-

- ACEITUNO, P., 1977. "Variaciones climáticas en el Norte Chico". Primer Seminario Nacional sobre desertificación, Santiago, Julio 1977.
- AKAIKE, H., 1970. "Statistical predictor identification, Ann. Inst. Stat. Math. 22, 203-217.
- AKAIKE, H., 1974. "A new look at the statistical model identification. IEEE Trans. Automat. Control, AC-19, 716-723.
- ANDERSEN, N., 1974. "On the calculation of filter coefficients for maximum entropy spectral analysis. Geophysics, 39, 1, 69-72.
- BURG, J.P., 1968. "A new analysis technique for time series data. Artículo presentado en el Advanced Study Institute on Signal Processing, OTAN, Enschede, Holanda.
- BURG, J.P., 1975. "Maximum entropy spectral analysis. Stanford Univ. Ph. D. Dissertation.
- HENRICI, P., 1974. "Elements of numerical analysis. Wiley, New York.
- HOLTON, J.R., 1975. "The dynamic meteorology of the stratosphere and mesosphere. Meteor. Monographs, 15, 37, American Meteor. Soc., Boston.
- JOHNSON, S.J. y N. ANDERSEN, 1978. "On power estimation in maximum entropy spectral analysis. Geophysics, 43, 4, 681-690.
- LACOSS, R.T., 1971. "Data adaptive spectral analysis methods. Geophysics, 36, 4, 661-675.
- LEVINSON, H., 1947. "The Wiener RMS error criterion in filter design and prediction. J. Math. Phys. 25, 261-278.
- MASON, B.J., 1976. "Towards the understanding and prediction of climate variations. Quart. Jour. Royal Meteor. Soc., 102, 433, 473-498.
- MENZEL, D.H., 1959. "Our sun". Harvard University Press, Cambridge, Massachusetts.

- SMYLIE, D.E., G.K. CLARKE y T.J. ULRICH, 1973. "Analysis of irregularities in the earth's rotation, en *Methods in Computational Physics*, 13, 391-430; Academic, New York.
- TABONY, R.C., 1979. "A spectral and filter analysis of long-period rainfall records in England and Wales. *The Meteorological Magazine*, 108, 1281, 97-118.
- TREITEL, S., P.R. GUTOWSKI y E.A. ROBINSON, 1977. "Empirical spectral analysis Revisited. En *Topics in Numerical Analysis III*, Academic Press.
- TRENBERTH, K.E., 1975. "A quasi-biennial standing wave in the Southern Hemisphere and interrelations with sea surface temperature, *Quart. Jour. Royal Meteorological Soc.*, 101, pp. 55-74.
- TRENBERTH, K.E., 1976. "Fluctuations and trends in indices of the Southern Hemisphere circulation. *Quart. Jour. Royal Meteor. Soc.*, 102, 431, 65-76.
- ULRICH, T.J. y T.N. BISHOP, 1975. Maximum entropy spectral analysis and autoregressive decomposition. *Rev. in Geoph. and Space Phys.*, 13, 1, 183-200.
- ULRICH, T.J. y R.W. CLAYTON, 1978. "Time series modelling and maximum entropy. *Phys. Earth Planet. Inst.*, 12, 188-200.
- WRIGHT, P.B., 1968. "A widespread biennial oscillation in the troposphere. *Weather*, 23, 2, 50-54.