

MODELOS AUTORREGRESIVOS INTEGRADOS CON PROMEDIO MOVIL
Y SU APLICACION A OBSERVACIONES DE TEMPERATURA
AMBIENTE DE LA ZONA N. OESTE DE MEXICO.

por:

Hugo G. Peña
Centro de Investigación Científica y
de Educación Superior de Ensenada, B.C.
Espinoza No. 843
Ensenada, Baja California, México.

RESUMEN

A partir de los modelos propuestos por G.E.P. Box y G. Jenkins se construye un modelo estocástico lineal estacional que se ajusta a datos observados de temperatura ambiente. El primer paso en la formulación de tal modelo es identificar la estructura del modelo candidato estimando la función de autocorrelación (acf), la función de autocorrelación parcial (pacf) y la función de autocorrelación inversa (iacf). Este trabajo describe la teoría de los modelos multiplicativos autoregresivos integrados con promedio móvil (ARIMA) y los pasos para la identificación del modelo particular de la familia, la estimación de los parámetros y las pruebas estadísticas de su ajuste. Se da un ejemplo de la predicción de la temperatura. El modelo ajustado a estos datos explica la variabilidad estacional de las medias mensuales pero no explica la variabilidad estacional de las desviaciones típicas mensuales. Por esta razón, este modelo es algo limitado. La metodología es extendida fácilmente a otro tipo de problemas.

ABSTRACT

Stochastic Seasonal Linear Model, using historical time serie data for temperature, is built by means of parametric time series models proposed by G.E.P. Box and G. Jenkins.

The first step in building such a model is to identify a candidate model structure by estimating the autocorrelation function (acf), partial autocorrelation function (pacf) and inverse autocorrelation function (iacf).

This paper describes the theory of multiplicative seasonal autoregressive integrated moving average (ARIMA) models and steps for identifying the particular member of the family, estimating the parameters and checking the fit. An illustration of forecast for temperature data is given. The seasonal ARIMA model fitted accounts for the seasonal variability of the monthly means but not the seasonal variability of the monthly standard deviation. For this reason its value is limited. The method is easily extended to other problems.

I INTRODUCCION

Desde la antigua Grecia, existe la controversia de dos escuelas en el estudio de los fenómenos naturales: la determinística y la estocástica. Desde entonces ambas tendencias han ido reemplazándose o complementándose mutuamente.

Es ampliamente conocido el hecho de que los fenómenos naturales contienen cierto tipo de información que impide el poder conocer su comportamiento de una manera exacta, a partir del único conocimiento de su estado presente. Aunque pueden poseer ciertas componentes determinísticas (debidas generalmente a fenómenos astronómicos), dichos procesos contienen, en algún grado, contribuciones no-determinísticas.

Es este tipo de contribución no-determinística el que es interesante de estudiar y el que motiva la realización de este trabajo.

A fin de efectuar diseños adecuados de ciertos sistemas (físicos, biológicos, económicos, etc.) es necesario conocer la estructura estadística de las variables que están involucradas. Al mismo tiempo, el conocimiento del carácter estadístico de las variables permitirá por ejemplo, realizar "simulaciones" del sistema con la intención de poder prever ciertas condiciones extremas a las que el sistema podrá estar expuesto. Otro aspecto no menos importante que surge del comportamiento estadístico de un sistema estriba en las posibilidades de su "predicción" a tiempos futuros.

Debido a estas consideraciones, este trabajo ilustra una relativamente nueva metodología basada en una familia de modelos lineales, con su aplicación a datos observados de temperaturas.

La metodología en la construcción de estos modelos estocásticos requiere de tres etapas importantes que se describen en la sección II, luego de una breve introducción a sus definiciones y nomenclatura. A través del trabajo se analiza la importancia de las funciones de autocorrelación (acf), de autocorrelación parcial (pacf) y de autocorrelación inversa (iacf) en la identificación adecuada del modelo. También se describen las pruebas estadísticas necesarias para consolidar el modelo.

MODELOS AUTORREGRESIVOS INTEGRADOS

En la sección V se muestra el aspecto "predictivo" del modelo adoptado.

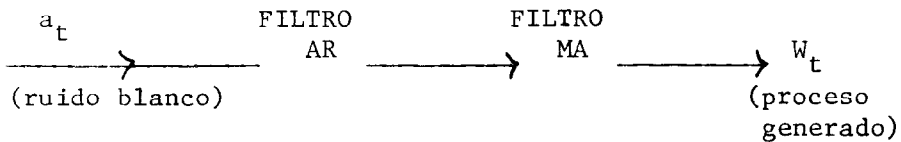
En general puede decirse que la potencialidad de este tipo de modelos reside en el hecho de que para su representación el requerimiento de la cantidad de parámetros a estimar es mínima, aunque haya que pagar "cierto precio" por ello.

II TEORIA Y METODO DE ANALISIS

En el año 1970, Box y Jenkins publicaron su libro "TIME SERIES ANALYSIS, FORECASTING AND CONTROL". En él se describen una familia de modelos paramétricos lineales llamados AUTOREGRESSIVE INTEGRATED MOVING AVERAGE, los cuales y por brevedad, llamaremos ARIMA.

Un proceso ARIMA puede ser pensado como aquel que es generado a partir del filtrado lineal de un proceso de ruido puro (ruido "blanco") que posee una distribución gaussiana con media nula y variancia σ_a^2

Esquemáticamente se puede representar como



El problema de seleccionar el modelo adecuado se reduce entonces a conocer las funciones de transferencia de ambos filtros (AR: autoregresivo, MA: promedio móvil).

La construcción de un modelo ARIMA requiere del desarrollo de 3 etapas importantes;

a) identificación de la cantidad de parámetros que llevará el modelo; b) estimación de los parámetros y c) aplicación de pruebas de diagnóstico a fin de verificar si el modelo ya identificado y estimado es el adecuado.

II.1 DEFINICIONES Y NOMENCLATURA

Sea Z_1, Z_2, \dots, Z_n una serie temporal evaluada a intervalos discretos e iguales.

La familia de los modelos ARIMA multiplicativos (también llamados ARIMA ESTACIONALES) se representa.

$$(B) \phi(B^s) (1-B)^d (1-B^s) D_{Z_t} = \theta(B) H(B^s) a_t \quad (1)$$

donde los símbolos significan

s: período estacional

B: operador desplazamiento definido tal que

$$B^{\ell} Z_t = Z_{t-\ell} \quad (\ell \text{ representan la variable "tiempo" en forma discreta})$$

a_t : Ruido blanco $N(0, \sigma_a^2)$

$\phi(B)$: operador autoregresivo (AR) no-estacional

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_P B^P \quad (\text{de orden } P) \quad (2)$$

$(1-B)^d \equiv \nabla^d$: operador diferencia de orden d no-estacional

$\phi(B^S)$: operador autoregresivo estacional (SAR)

$$\phi(B^S) = 1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_P B^{PS} \quad (\text{orden } P) \quad (3)$$

$(1-B^S)^D \equiv \nabla_S^D$: operador diferencia de orden D estacional

$\theta(B)$: operador promedio móvil (MA) no-estacional.

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (\text{de orden } q) \quad (4)$$

$H(B^S)$: operador promedio móvil estacional (SMA)

$$H(B^S) = 1 - h_1 B^S - \dots - h_Q B^{QS} \quad (\text{de orden } Q) \quad (5)$$

La notación para representar un modelo ARIMA multiplicativo es: $(p,d,q) \times (P,D,Q)_S$. El grupo del primer paréntesis representa la forma del modelo en su parte no-estacional y el segundo grupo es el análogo en su parte estacional.

II.2 METODO DE ANALISIS

La forma de operar se puede resumir esquemáticamente

DATOS HISTORICOS

1. Graficado de las series $Z_t, \nabla Z_t, \nabla_S Z_t$, etc.

2. Funciones de Autocorrelación (acf)

IDENTIFICACION 3. Funciones de Autocorrelación Parcial (pacf)

4. Funciones de Autocorrelación Inversa (iacf)

5. Periodograma acumulativo

ESTIMACION 1. Estimación por MAXIMA VEROSIMILITUD (MAXIMUM LIKELIHOOD) de los parámetros

2. Transformaciones no-lineales de BOX-COX

1. Sobreajuste (over-fitting)

2. Examinación de la estructura estadística de los residuos para comprobación de su normalidad e independencia.

PRUEBAS DE DIAGNOSTICO 3. Periodograma acumulativo como prueba de ruido blanco.

4. Asimetría y kurtosis de los residuos.

MODELOS AUTORREGRESIVOS INTEGRADOS

5. Prueba para observar cambios en la varianza de los residuos.

El propósito de la etapa de identificación es determinar el grado de diferenciación requerida para lograr la estacionaridad de la serie y para determinar el orden de los operadores AR-MA estacionales y no estacionales. El comportamiento de las funciones acf, pacf e iacf dependen (en cuanto a sus "formas" y en cuanto a sus valores numéricos) de los ordenes de los operadores (Box y Jenkins, 1970).

En la etapa de estimación, se realiza la estimación numérica de los parámetros, empleando el método de la "suma incondicional de los cuadrados de los residuos". Minimizando esta función respecto de los parámetros se obtiene una estimación de ellos, la cual es óptima en el sentido de mínimos cuadrados. Esta estimación corresponde a la de máxima verosimilitud cuando la distribución de densidad de probabilidad de los residuos que resultan del ajuste es gaussiana. Para efectuar la minimización de la suma de los cuadrados existen varios algoritmos. En este trabajo se usa el algoritmo de Marquardt (1963).

En la etapa de diagnóstico se realizan pruebas estadísticas para verificar la normalidad e independencia de los residuos. Las pruebas efectuadas en este trabajo son: a) estudio de la función de autocorrelación de los residuos; b) histograma de los residuos; c) estimación del estadístico Q (Box y Jenkins, 1970); d) periodograma acumulativo de los residuos.

La verificación de la independencia de los residuos es la más crítica. La violación de este requerimiento invalida el modelo ajustado y es necesario comenzar otra vez por la etapa de identificación. Los requisitos de normalidad y estacionaridad en la varianza de los residuos son menos críticos y la violación de ellos puede ser removida mediante una transformación no-lineal adecuada de los datos originales (Box y Cox, 1964).

III DATOS UTILIZADOS

Los datos a modelar corresponden a observaciones de temperatura ambiente. Son valores promedios mensuales de la zona de San Diego, California, USA, fueron extraídos del CLIMATIC SUMMARY OF THE UNITED STATES. Se eligió esta estación por ser la más cercana a la Zona Norte de Baja California, México. Los datos comprenden el intervalo entre 1912 y 1970. Se encuentran en preparación los modelos ARIMA para las temperaturas extremas (valores mensuales máximo y mínimo).

IV RESULTADOS Y DISCUSION

IV.1 IDENTIFICACION DEL MODELO

Se estiman las funciones de autocorrelación (acf) para los siguientes esquemas de diferenciación: (d=1, D=0), (d=0, D=1) y (d=1, D=1). Sus respectivas gráficas se muestran en la fig. 1 a, b, y c. Una inspección de esta nos indica que un posible esquema estacionario es el que corresponde a (d=1, D=1) (Fig. 1 c), ya que su acf ha decaído a cero mucho más rápidamente que en los demás esquemas.

Esto nos conduce a un modelo multiplicativo (0, 1, 1)_x (0, 1, 1)₁₂. La acf(k) teórica de tal modelo posee todos los acf(k) nulos, excepto para k=1, 11, 12 y 13, y además requiere que acf(11)=acf(13). La fig. 1c muestra justamente estas características.

IV.2 ESTIMACION DEL MODELO (0, 1, 1)_x(0, 1, 1)₁₂

El modelo propuesto se puede describir, de acuerdo a (1) y teniendo en cuenta que p=0, d=1, q=1, P=0, D=1, Q=1, S=12, como:

$$(1-B) (1-B^{12}) Z_t = (1-\theta_1 B) (1-h_1 B^{12}) a_t \quad (6)$$

La acf teórica de este modelo se obtiene considerando que

$$W_t \equiv (1-B)(1-B^{12})Z_t = \psi(B) a_t$$

donde el operador $\psi(B)$ es

$$\psi(B) = 1-\theta_1 B - h_1 B^{12} + \theta_1 h_1 B^{13}$$

y por lo tanto

$$\psi_0=1, \psi_1 = -\theta_1, \psi_{12} = -h_1, \psi_{13} = \theta_1 h_1, \psi_i = 0 \quad (7)$$

siendo todos los restantes $\psi_i = 0$

Se puede demostrar que, conocidos los elementos ψ_i del operador $\psi(B)$, la función de autocorrelación es

$$acf(k) = \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+k} \right) / \left(\sum_{i=0}^{\infty} \psi_i^2 \right) \quad (8)$$

Haciendo uso de los ψ_i hallados (7) y de (8), resulta

$$acf(1) = -\theta_1 / (1+\theta_1^2)$$

$$acf(11) = acf(13) = (\theta_1 h_1) / \{ (1+\theta_1^2)(1+h_1^2) \}$$

$$acf(12) = -h_1 / (1+h_1^2)$$

con todos los restantes acf's nulos.

La estimación de los parámetros θ_1 y h_1 usando el algoritmo de Marquardt es

$$\hat{\theta} = 0.7024 \pm 0.0270 \quad \text{y} \quad \hat{h} = 0.09554 \pm 0.0127$$

MODELOS AUTORREGRESIVOS INTEGRADOS

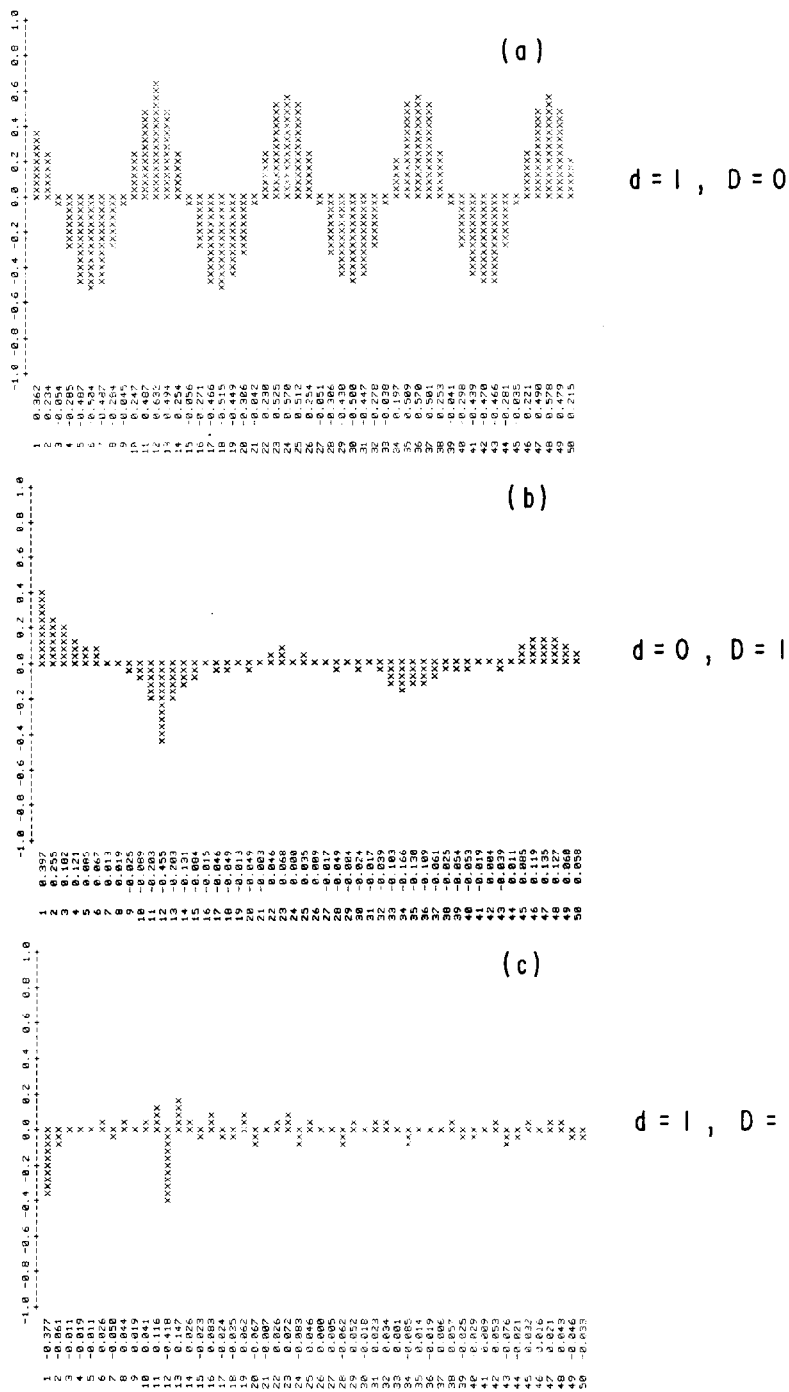


FIG.1 FUNCIONES DE AUTOCORRELACION OBSERVADAS PARA LOS ESQUEMAS INDICADOS.

La desviación típica de \hat{h}_1 indica que, aunque \hat{h}_1 está sospechosamente cercano a la unidad, la estimación es significativamente diferente de 1.

IV.3 PRUEBAS DE DIAGNOSTICO

La función de autocorrelación de los residuos (racf) del modelo se ilustra en la fig. 2 y el histograma en la fig. 3.

El histograma muestra una razonable aproximación a una curva gaussiana. La racf se aparta sospechosamente de la correspondiente a "ruido blanco". En efecto, no solo existen varias estimaciones significativamente diferentes de cero ($\sigma \sim \frac{1}{\sqrt{n}} = 0.038$), sino que existe una marcada tendencia a poseer valores negativos, situación esta impropia para el comportamiento de una serie puramente aleatoria.

La prueba del estadístico Q (Box y Jenkins, 1970) es altamente significativa, ratificando así la conclusión de que el residuo se aparta de la independencia estadística requerida para validar el modelo. Sin embargo, ha de notarse que la componente estacional (período anual $S=12$) en la racf ha sido casi completamente eliminada, ya que desde el retardo $k=10$ hasta el $k=15$, la racf es menor que la desviación típica.

Estos resultados muestran que aunque la parte estacional del modelo parece haber sido identificada, no ha sucedido lo mismo con la parte no-estacional. Evidentemente ha de intentarse una nueva identificación en busca de otro modelo.

IV.4 NUEVA IDENTIFICACION

Efectuando un análisis más detallado se observa (Fig. 1.6) que la acf muestra un comportamiento aproximadamente exponencial decreciente en los primeros retardos, sugiriendo la presencia de una componente AR no-estacional. El valor altamente significativo en el retardo $k=12$ y la estimación casi nula a múltiplos de 12 sugiere la existencia de una componente MA estacional. A fin de establecer estos hechos, se han graficado las pacf (Fig. 4a) e iacf (Fig. 5a) del esquema ($d=0, D=1$). La pacf muestra claramente las características de un modelo multiplicativo $(2,0,0) \times (0,1,1)_{12}$ debido a que se trunca después del segundo retardo, y a retardos múltiplos de 12 se observa una atenuación que indica la existencia de una componente MA. La iacf (Fig. 5a) corrobora esta identificación dando más validez al modelo $(2,0,0) \times (0,1,1)_{12}$:

$$(1 - \phi_1 B - \phi_2 B^2) w_t = (1 - h_1 B^{12}) a_t \quad (9)$$

donde $w_t = (1 - B^{12}) z_t$

MODELOS AUTORREGRESIVOS INTEGRADOS

FIG. 2 ACF DEL RESIDUO DEL MODELO MULTIPLICATIVO $(0,1,1) \times (0,1,1)_{12}$

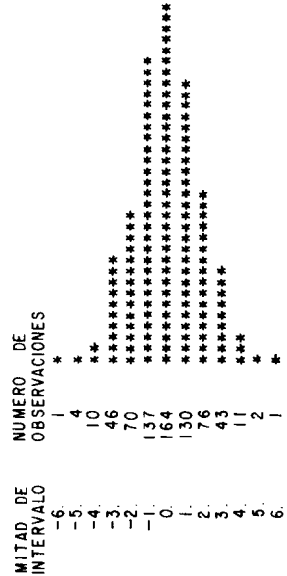
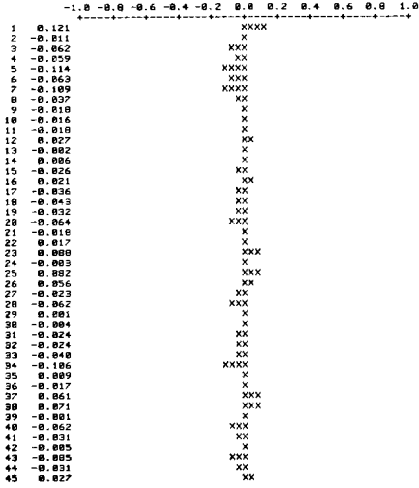


FIG. 3 HISTOGRAMA DEL RESIDUO DEL MODELO MULTIPLICATIVO $(0,1,1) \times (0,1,1)_{12}$

FIG. 4a FUNCION DE AUTOCORRELACION PARCIAL OBSERVADA PARA EL ESQUEMA $(d=0, D=1)$

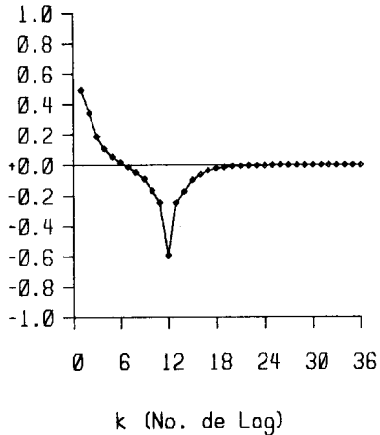
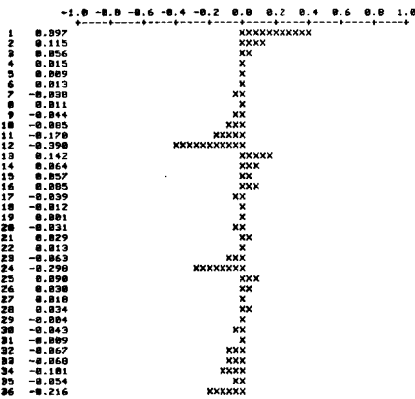


FIG. 4b Funcion de autocorrelacion teorica para el esquema $(d=0, D=1)$ y el modelo $(2,0,0) \times (0,1,1)_{12}$.

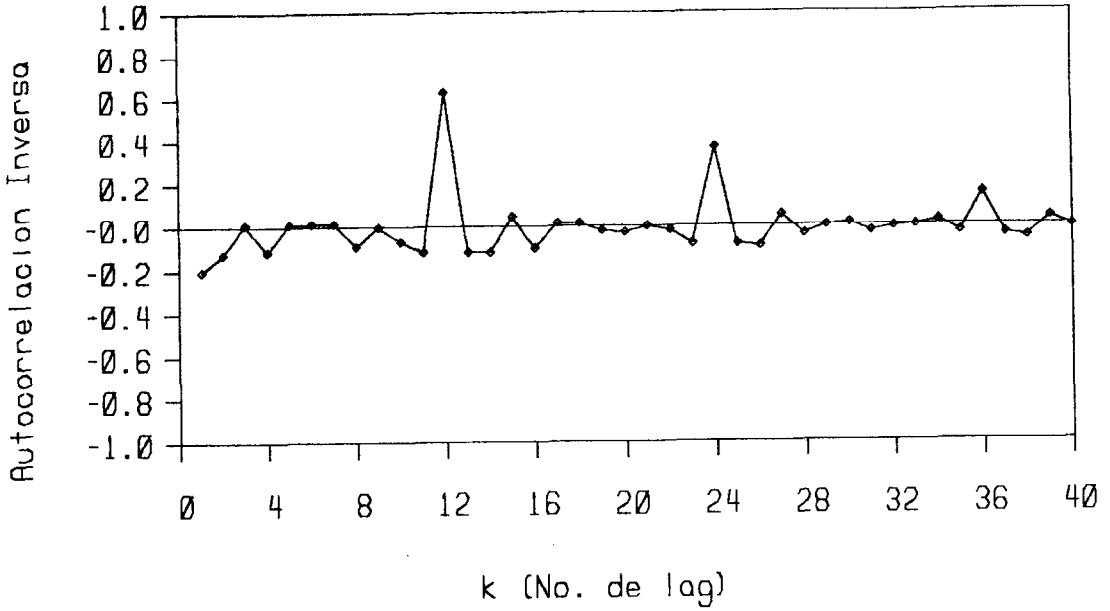


FIG.5a Función de autocorrelación Inversa observada para el esquema ($d=0.0=1$).

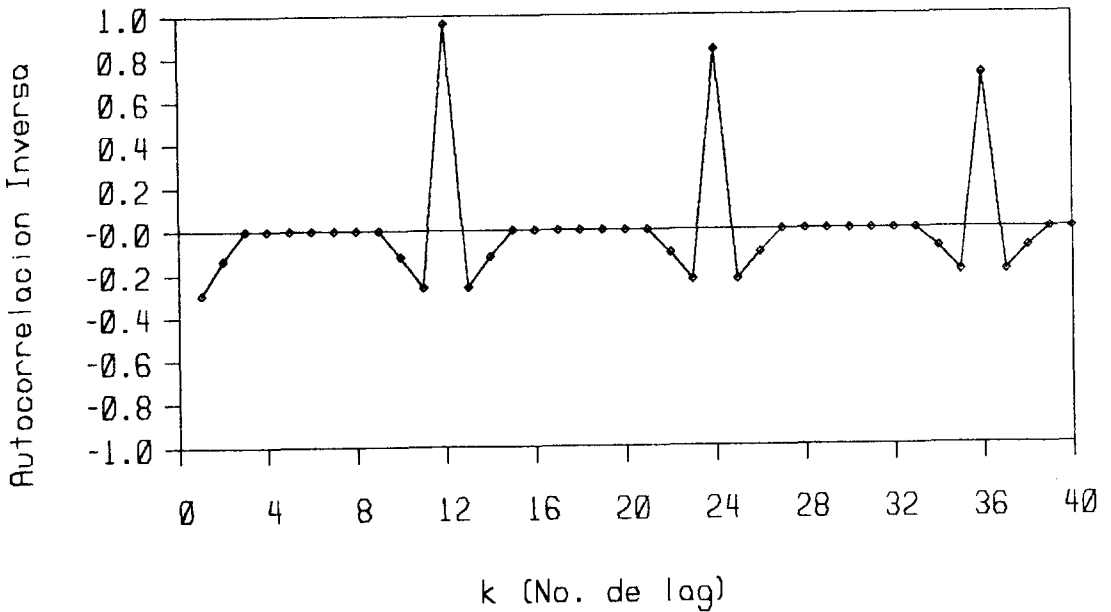


FIG.5b Función de autocorrelación Inversa teórica para el esquema ($d=0.0=1$) y modelo ARIMA $(2.0.0) \times (0.1.1)_{12}$

MODELOS AUTORREGRESIVOS INTEGRADOS

IV.5 ESTIMACION DEL MODELO $(2,0,0) \times (0,1,1)_{12}$

Los resultados de la estimación iterativa de los parámetros ϕ_1 , ϕ_2 y h_1 por medio del algoritmo de Marquardt se presentan en la Tabla I. Con fines comparativos, las figs. 4b y 5b muestran las acf e iacf teóricas, suponiendo como exactos los valores estimados de los parámetros.

Estas funciones teóricas se obtienen de la siguiente forma: un modelo ARMA se representa como $R(B)w_t = S(B)a_t$, donde el operador $R(B)$ contiene toda la información autorregresiva y el $S(B)$ contiene toda la información promedio móvil. Luego, podemos expresar el modelo de dos formas diferentes.

como MA puro

$$w_t = S(B)R^{-1}(B)a_t \equiv \psi(B)a_t \quad (10)$$

y como AR puro

$$S^{-1}(B)R(B)w_t \equiv T(B)w_t = a_t \quad (11)$$

en (10) es: $\psi(B) \cdot R(B) = S(B) \quad (12)$

y en (11) es: $T(B) \cdot S(B) = R(B) \quad (13)$

igualando coeficientes en cada una de las dos expresiones (12) y (13), se obtienen los coeficientes de los operadores $\psi(B)$ y $T(B)$. Con ellos se obtienen las acf e iacf respectivamente.

TABLA I. ESTIMACION DE LOS PARAMETROS DEL MODELO $(2,0,0) \times (0,1,1)_{12}$ POR EL ALGORITMO DE MARQUARDT (1963) (59 AÑOS)

ITERACION	PARAMETRO		
	ϕ_1	ϕ_2	θ_1
1	0.169	0.115	0.250
2	0.222	0.128	0.400
3	0.264	0.139	0.550
4	0.297	0.147	0.700
5	0.324	0.151	0.850
6	0.349	0.147	0.957
7	0.360	0.144	0.959
8	0.360	0.144	0.959

ESTIMACION FINAL DE LOS PARAMETROS

	ESTIMACION	DESVIACION ST.
ϕ_1	0.3599	0.0376
ϕ_2	0.1442	0.0377
θ_1	0.9592	0.0135
σ_a^2	2.88	--

$$\text{acf}(k) = \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+k} \right) / \left(\sum_{i=0}^{\infty} \psi_i^2 \right) \quad (14)$$

$$\text{iacf}(k) = \left(\sum_{i=0}^{\infty} T_i T_{i+k} \right) / \left(\sum_{i=0}^{\infty} T_i^2 \right) \quad (15)$$

Se diseñó un programa de computación en FORTRAN IV para hallar los coeficientes ψ_i y T_i a partir del conocimiento de los coeficientes R_i y S_i , complementándolo con una subrutina que efectúa la operación de convolución indicada en (14) y (15).

Como se observa en las Figs. 1.6 y 4b, las acf observada y teórica corresponden razonablemente. Lo mismo sucede con las iacf representadas en las figs. 5a y 5b. Esto conduce una vez más a la ratificación del modelo multiplicativo $(2,0,0) \times (0,1,1)_{12}$ como el mejor identificado.

IV.6 PRUEBAS DE DIAGNÓSTICO DEL MODELO $(2,0,0) \times (0,1,1)_{12}$

La racf de los residuos de este modelo se ilustra en la Fig. 6. En principio, y por la apariencia de esta función, no hay razón para sospechar que no proviniese de una serie puramente aleatoria. Pero desde luego, esto no es suficiente. El estadístico Q calculado para los primeros 36 retardos es 37.1. El valor de X^2 al nivel de significación del 10% para 33 grados de libertad es 43. Comparando estos valores, vemos que $Q < X^2_{10\%}$ (10%), por lo tanto los residuos pueden ser aceptados como aleatorios puros. Para ratificar esta conclusión se construyó el periodograma acumulativo de los residuos (Fig. 7) el cual no muestra desviaciones significativas del carácter aleatorio de los mismos. El histograma de los residuos, evaluando su asimetría y Kurtosis, no difiere significativamente del correspondiente a una distribución normal, lo cual refuerza una vez más el hecho de que los residuos se comporten estadísticamente independientes y con distribución gaussiana.

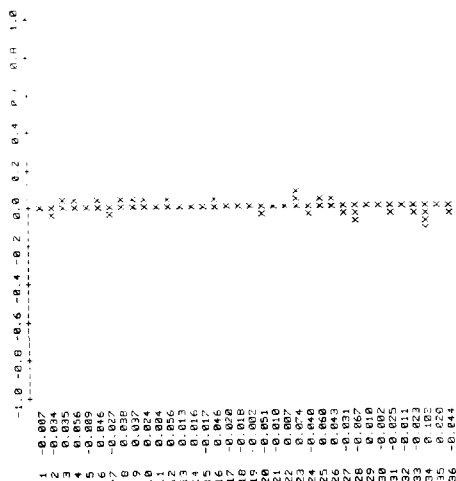


FIG. 6 FUNCION DE AUTOCORRELACION DEL RESIDUO DEL MODELO MULTIPLICATIVO $(2,0,0) \times (0,1,1)_{12}$

MODELOS AUTORREGRESIVOS INTEGRADOS

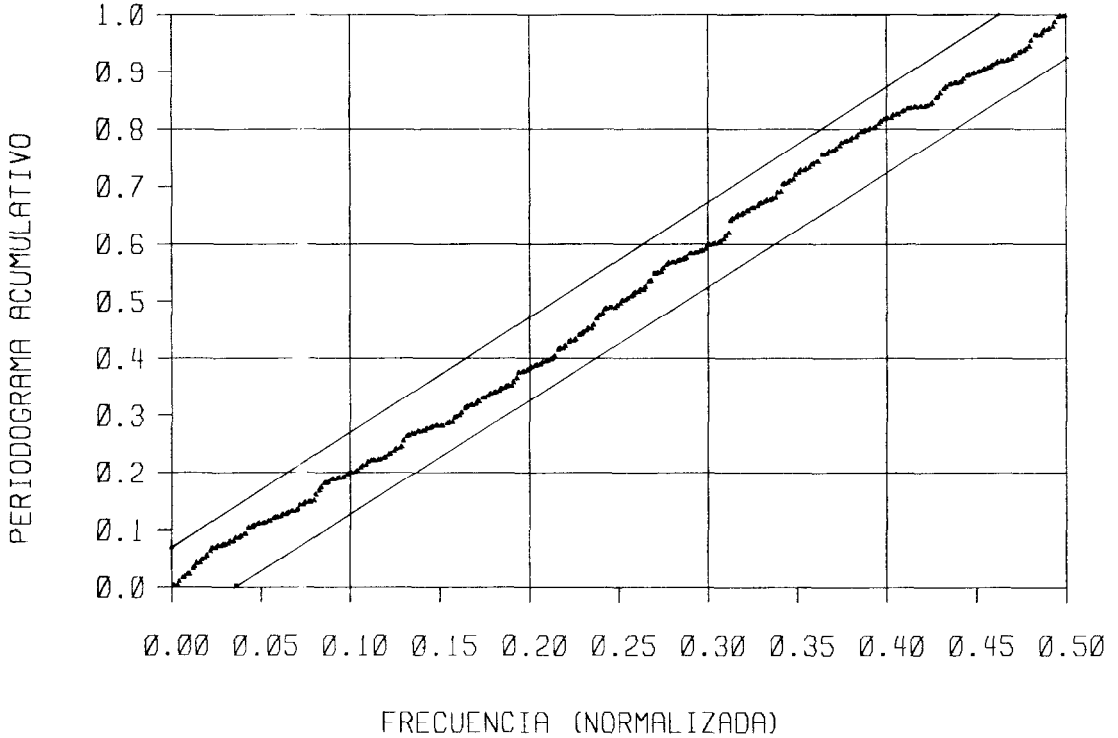


FIG.7 PERIODOGRAMA ACUMULATIVO DE FRECUENCIA (NORMALIZADO)
DEL RESIDUO DEL MODELO MULTIPLICATIVO $(2.0,0) \times (0.1,1)_{12}$.
LIMITES DE CONFIANZA AL 95%.

Habiendo los residuos pasado las pruebas más significativas, se concluye que el modelo identificado como el $(2,0,0) \times (0,1,1)_{12}$ es el que mejor se ajusta a los datos históricos.

V PREDICCIÓN

Box y Jenkins (1970) muestran que la predicción a tiempo futuro ℓ , con origen t , que posee el mínimo error cuadrático medio es el valor medio condicional dado por

$$\langle Z_{t+\ell} \rangle \equiv \hat{Z}_t(\ell) = E(Z_{t+\ell} / Z_t, Z_{t-1}, \dots, Z_1)$$

donde el símbolo $E(\cdot)$ expresa la esperanza matemática. El error en la estimación de la predicción se distribuye normalmente con media nula y varianza.

$$\text{Var}\{e_t(\ell)\} = \sigma_a^2 \left(\sum_{k=0}^{\ell-1} \psi_k^2 \right) \quad (16)$$

siendo ψ_k los coeficientes del operador que cumple la relación

$$Z_t = \psi(B)a_t = \sum_{k=0}^{\infty} \psi_k a_{t-k}$$

De esta forma, los límites de confianza al nivel $(1-\alpha)100\%$ estarán dados por

$$\hat{Z}_t(\ell) \pm U_{\alpha/2} \left(\sum_{k=0}^{\ell-1} \psi_k^2 \right)^{1/2} \sigma_a \quad (17)$$

donde $U_{\alpha/2} = 1.96$, para $\alpha=0.05$ (5% del nivel de significación). Los coeficientes ψ_k se obtienen con el programa antes mencionado, a partir del modelo ya adoptado $(2,0,0) \times (0,1,1)_{12}$ que satisface la relación

$$\psi(B) \cdot (1-\phi_1 B - \phi_2 B^2) \cdot (1-B^{12}) = (1-h_1 B^{12})$$

V.1 LA FUNCION DE PREDICCIÓN

Para efectuar la predicción expresamos el modelo propuesto, en su forma de ecuación de diferencias y evaluado en $(t+\ell)$,

$$(1-\phi_1 B - \phi_2 B^2)(1-B^{12}) \cdot Z_{t+\ell} = (1-h_1 B^{12}) a_{t+\ell}$$

Desarrollando esta expresión

$$Z_{t+\ell} = \phi_1 Z_{t+\ell-1} + \phi_2 Z_{t+\ell-2} + Z_{t+\ell-12} - \phi_1 Z_{t+\ell-13} - \phi_2 Z_{t+\ell-14} + a_{t+\ell} - h_1 a_{t+\ell-12} \quad (18)$$

tomando la esperanza matemática a (18), y teniendo en cuenta que se satisfacen las siguientes relaciones

$$E(Z_{t-k}) = Z_{t-k}, \quad k=0,1,2,\dots$$

$$E(Z_{t+k}) = \hat{Z}_t(k), \quad k=1,2,\dots$$

$$E(a_{t-k}) = a_{t-k} = Z_{t-k} - \hat{Z}_{t-k-1}(1), \quad k=0,1,2,\dots$$

$$\text{y } E(a_{t+k}) = 0, \quad k=1,2,\dots$$

obtenemos la función de predicción con origen t que puede ser evaluada para $\ell=1,2,3,\dots$ etc.

Por ejemplo, para $\ell=1$, es

$$\hat{Z}_t(1) = \phi_1 Z_t + \phi_2 Z_{t-1} + Z_{t-11} - \phi_1 Z_{t-12} - \phi_2 Z_{t-13} - h_1 a_{t-11}$$

A partir de un cierto valor de ℓ mayor que el orden del operador promedio móvil, la predicción sólo depende del operador autorregresivo, mientras que el error de la predicción depende de ambos operadores y de la varianza del residuo σ_a^2 .

MODELOS AUTORREGRESIVOS INTEGRADOS

V.2 RESULTADOS Y DISCUSION DE LA PREDICCION

Para efectuar la predicción y compararla con datos reales, se adoptó el siguiente esquema: se acortó la serie original de 59 años quitando los últimos 4 años, y con los 55 años restantes se recalcularon los parámetros del modelo. De esta forma se pueden comparar los valores observados y predichos de los últimos 4 años. La Tabla II muestra la nueva estimación de los parámetros. Si se compara con la Tabla I se ve que no existen diferencias significativas entre ambas estimaciones. Esto indica una aceptable estabilidad en los datos observados.

En la Tabla III se muestran los valores medios de la temperatura para cada mes del año, junto con su desviación típica. La Tabla IV presenta los valores numéricos del primero de los 4 años de predicción, con sus correspondientes intervalos de confianza al 95%. La Fig. 8 muestra las temperaturas predichas y observadas para los 4 años de predicción. Como se puede notar en la Fig. 8, la predicción se compara favorablemente con las observadas. Si comparamos las temperaturas predichas con las medias mensuales de todos los años (Tabla V) vemos que la diferencia afecta sólo a los décimos de grados °F. Este resultado era de esperarse ya que para una serie libre de tendencias, la predicción tiende a la estimación incondicional de los valores medios.

La Fig. 9 compara el error de la predicción con las desviaciones típicas mensuales (estas últimas se repiten cíclicamente cada 12 meses). Este resultado es importante ya que demuestra la desventaja de todo modelo multiplicativo. Debido a su construcción, el modelo multiplicativo es incapaz de modelar la variación de las desviaciones típicas. La diferenciación estacional ($D=1$) ha eliminado dicha variabilidad. Es el precio que se paga por la reducción del número de parámetros que han de representar este tipo de modelos.

Finalmente, la Fig. 10 muestra el ajuste del modelo durante 150 meses a partir de enero de 1913. Se observa que dicho ajuste es excelente.

VI RESUMEN Y CONCLUSION

Se ha ilustrado como los modelos ARIMA pueden ser adoptados para modelar procesos naturales que poseen componentes estacionales. Su simplicidad es inmediata si se toma en cuenta la pequeña cantidad de parámetros que se necesitan para su representación. Si el modelo es multiplicativo con $D \neq 0$, se pierde cierto tipo de información como sucedió con la variabilidad en las dispersiones típicas mensuales. Otros ti-

PEÑA

TABLA II. ESTIMACION DE LOS PARAMETROS DEL MODELO $(2,0,0) \times (0,1,1)_{12}$ POR EL ALGORITMO DE MARQUARDT (1963). (55 AÑOS)

ITERACION	PARAMETRO		
	ϕ_1	ϕ_2	θ_1
1	0.170	0.112	0.250
2	0.224	0.124	0.400
3	0.266	0.133	0.550
4	0.299	0.139	0.700
5	0.326	0.142	0.850
6	0.350	0.137	0.963
7	0.361	0.132	0.962
8	0.362	0.132	0.963
9	0.362	0.132	0.963

ESTIMACION FINAL DE LOS PARAMETROS.

	ESTIMACION	DESVIACION ST.
ϕ_1	0.3616	0.0391
ϕ_2	0.1318	0.0394
θ_1	0.9626	0.0144
σ_a^2	2.94	—

TABLA III. MEDIAS MENSUALES Y SUS DESVIACIONES ESTANDARD

	MEDIAS	DESVIACION ST.
ENERO	54.64	2.18
FEBRERO	55.32	2.24
MARZO	56.70	2.19
ABRIL	58.68	1.77
MAYO	60.87	1.39
JUNIO	63.95	1.57
JULIO	67.42	1.40
AGOSTO	68.70	1.74
SEPTIEMBRE	67.01	1.94
OCTUBRE	63.43	1.66
NOVIEMBRE	59.94	2.18
DICIEMBRE	56.24	2.05

MODELOS AUTORREGRESIVOS INTEGRADOS

TABLA IV. TEMPERATURAS (GRADOS F) OBSERVADAS Y PREDICHAS
(LIMITES DE CONFIANZA AL 95%).

PERIODO	PREDICCION	L. INFERIOR	L. SUPERIOR	OBSERVADAS
661	54.83	51.47	58.19	55.4
662	55.51	51.94	59.09	56.4
663	57.02	53.34	60.70	56.6
664	58.83	55.12	62.55	58.2
665	60.94	57.21	64.66	61.4
666	64.03	60.30	67.76	63.3
667	67.53	63.80	71.26	69.2
668	68.62	64.89	72.35	69.0
669	67.08	63.34	70.81	66.4
670	63.58	59.85	67.31	64.6
671	60.18	56.45	63.91	63.3
672	56.40	52.67	60.13	56.0

TABLA V. TEMPERATURAS PREDICHAS Y MEDIAS MENSUALES
(1er. AÑO DE PREDICCION. TEMPERATURA EN
GRADOS F)

PERIODO	PREDICCION	MEDIAS	DIFERENCIA
661(EN)	54.83	54.64	0.19
662(FE)	55.51	55.32	0.19
663(MZ)	57.02	56.70	0.32
664(AB)	58.83	58.68	0.15
665(MY)	60.94	60.87	0.07
666(JN)	64.03	63.95	0.08
667(JL)	67.53	67.42	0.11
668(AG)	68.62	68.70	-0.08
669(ST)	67.08	67.01	0.07
670(OC)	63.58	63.43	0.15
671(NO)	60.18	59.94	0.24
672(DI)	56.40	56.24	0.16

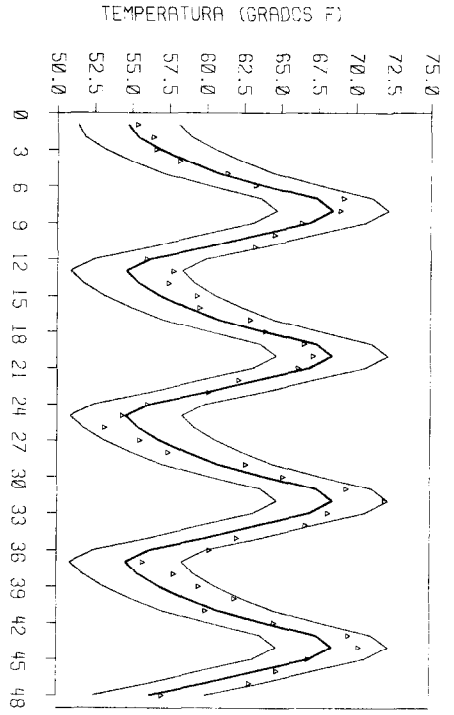


FIG. 8 TEMPERATURAS PREDICIDAS Y SUS INTERVALOS DE CONFIANZA AL 95 %
 LINEA SÓLIDA: TEMPERATURAS PREDICIDAS; LINEA FINA: INTERVALOS DE CONFIANZA.
 SÍMBOLO Δ: TEMPERATURAS OBSERVADAS.

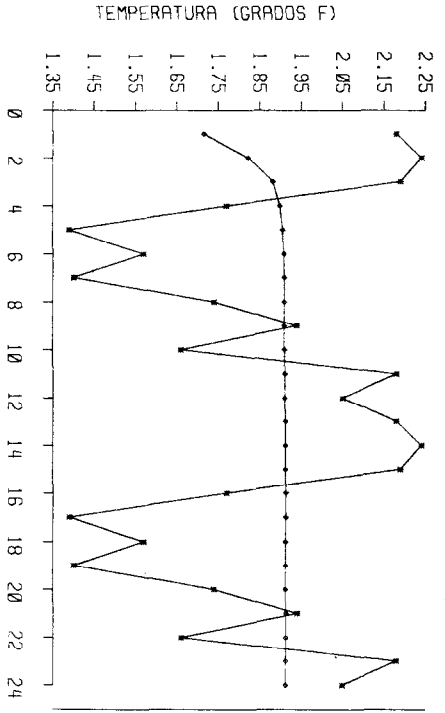


FIG. 9 ERROR ESTÁNDAR DE LA PREDICCIÓN (x) y DESVIACIONES STAD. DE LAS MEDIDAS (x*)

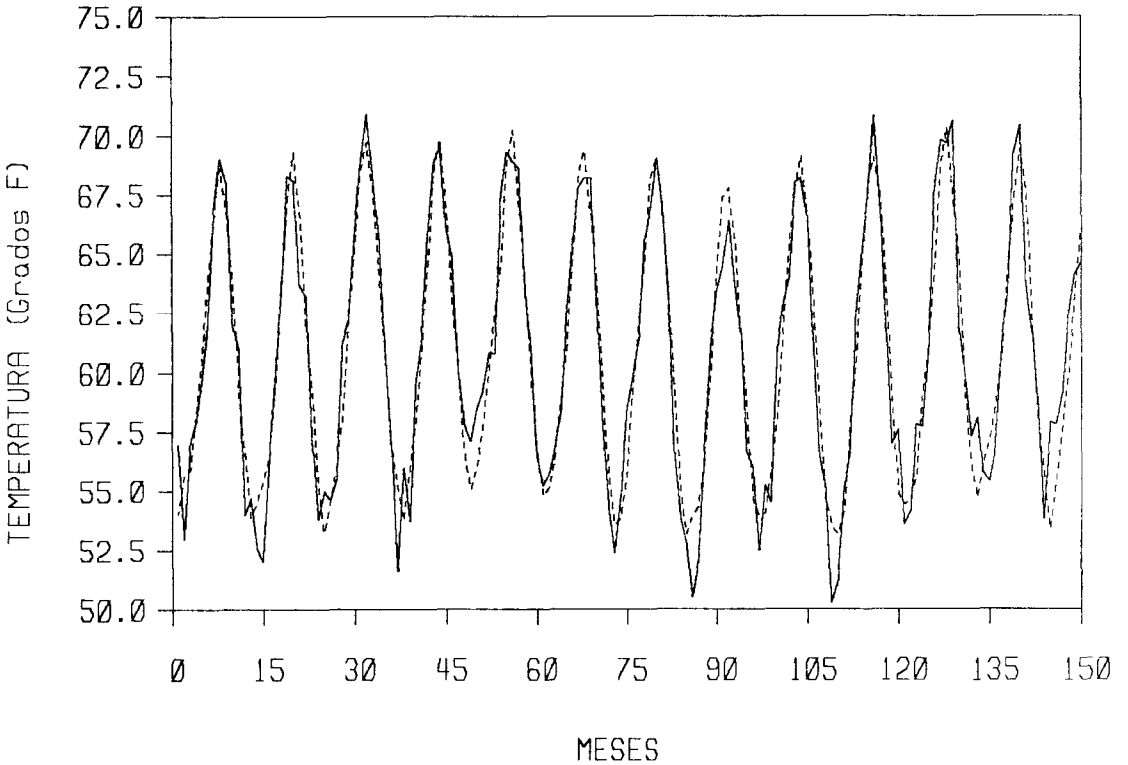


FIG. 10 AJUSTE DEL MODELO $(2,0,0) \times (0,1,1)_{12}$ A LOS DATOS OBSERVADOS DESDE ENERO 1913 HASTA JUNIO 1926.
 DATOS: LINEA LLENA; MODELO: LINEA QUEBRADA.

MODELOS AUTORREGRESIVOS INTEGRADOS

pos de modelos hubieran requerido la estimación de no menos de 24 parámetros (12 medias mensuales, 12 desviaciones típicas y 1, 2 ó 3 parámetros AR y/o MA no-estacionales).

Se mostró también como se procede en las etapas de identificación, estimación y de diagnóstico, a fin de obtener el mejor modelo que se adapte a los datos. En este aspecto se vió críticamente, como el único uso de la función de autocorrelación (acf) puede llevar a la identificación errónea del modelo. La inclusión de las funciones de autocorrelación parcial e inversa es un factor decisivo para el desarrollo de esta primer etapa.

En general puede decirse que la elección del modelo depende del propósito que se le quiera dar, es decir, si se lo requiere para la generación (simulación) o predicción, o para escalas temporales cortas o largas.

AGRADECIMIENTOS.

A la Señora Dolores Peralta Tubillo, por su paciente tarea de mecanografiar este artículo.

LITERATURA CITADA

BOX, G.E.P. y D.R. Cox, 1964. An analysis of transformations
J. Roy Statist. Soc. Ser. B., 26, 211-252.

BOX, G.E.P. y G.M. Jenkins, 1970. Time series analysis: Forecasting and Control. Holden-day, San Francisco, Calif.

CLEVELAND, W.S., 1972. The inverse autocorrelation of a time series and their applications. Technometrics, Vol. 12 No. 2 .

MARQUARDT, D.W., 1963. An algorithm for least squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math., 11(2):431-441.