

MODELOS ESTADÍSTICOS ESTRUCTURALES DE SERIES DE PROMEDIOS MENSUALES DE HELIOFANÍA PARA ALGUNAS LOCALIDADES DEL GRAN BUENOS AIRES

Orlando J. Avila Blas¹, Elizabeth Gisselle Collivadino¹ y Hugo Grossi Gallegos²

¹Av. Bolivia 5150 - (4400) Salta, Argentina. Tel. : (0387) 4255385,

Email : oblas@unsa.edu.ar, gisselle.collivadino@arnet.com.ar

²Rutas 5 y 7, (6700) Luján, Buenos Aires, Argentina, Fax : (02323) 425795, Email : grossi@mail.unlu.edu.ar

RESUMEN - En el presente trabajo realizamos un estudio estadístico de tres series de promedios mensuales de heliofanía, empleando la metodología del análisis estructural, correspondientes a datos tomados en las localidades de El Palomar, San Miguel y Don Torcuato. Hemos desarrollado la teoría matemática estadística específica a fin de presentar respectivos modelos, los que se pueden usar para poder generar sintéticamente valores de promedios mensuales de heliofanía. Esta modelización permite realizar pronósticos con muy alta confiabilidad (95%).

Palabras claves : heliofanía, promedios mensuales, estadística, modelos estructurales, pronósticos.

INTRODUCCION

Para poder realizar predicciones a largo plazo del comportamiento de los sistemas de aprovechamiento de la energía solar a menudo sólo es necesario tener conocimiento de los valores promedios mensuales de diferentes variables disponible en una localidad (por ejemplo, irradiación solar global en promedios mensuales, valores de heliofanía promedio mensual, etc.). Para otras aplicaciones, se hace necesario realizar una simulación más detallada para las que son necesarios datos diarios u horarios, y este tipo de datos son difíciles de obtener ó bien las series existentes tienen poco desarrollo temporal, a lo que se puede sumar presencia de datos atípicos (outliers) y grandes cambios estructurales. Este conjunto, por nombrar solo algunas, de situaciones en la descripción de los datos, dió origen a la búsqueda de alternativas que permitieran satisfacer la demanda de una mejor descripción estadística de los mismos.

El año pasado, se encontraron modelos estructurales de series de irradiación solar global promedio diarias, medidas en las ciudades de Córdoba, Marcos Juárez y Paraná (Avila Blas y Grossi Gallegos, 2002) con muy buenas propiedades estadísticas, lo cual se logró mediante la introducción de una transformación matemática especial denominada “índice de contribución relativa”, que permitió “levantar” la condición de normalidad en la distribución de los datos y atacar la presencia de datos atípicos o “outliers”. Últimamente se ha vuelto a prestar atención a la información de heliofanía teniendo en cuenta que la extensión de sus registros y el número de estaciones en las que se mide supera en mucho a las de irradiación solar. Es así que Suehrcke (2000) presentó una relación que permite estimar la irradiación global diaria a partir de los valores diarios de horas de insolación a través de la ecuación

$$(n/N) = (K/K_c)^3 \quad (1)$$

en donde n es el número de horas de sol registradas por un heliógrafo (heliofanía efectiva), N la longitud astronómica del día medida en horas, K el índice diario de claridad y K_c el índice de claridad correspondiente a un día claro, sugiriendo un valor de 0.73.

En el presente artículo hemos trabajado modelando series de promedios mensuales de heliofanía para las localidades de El Palomar, San Miguel y Don Torcuato obtenidos por el Servicio Meteorológico Nacional (S.M.N.) y facilitados por la Red Iberoamericana de Solarimetría (RISOL). Cada una de ellas tiene una longitud distinta y características especiales que impiden el

¹ Probabilidades y Estadística, Depto. de Matemática, Fac. Cs. Exactas - UNSa

² División Física, Universidad Nacional de Luján, Miembro de la Carrera de Investigador-CONICET

uso del índice antes mencionado a fin de poder aplicar el suavizado y filtrado de Kalman. De aquí la necesidad de desarrollar la teoría estadística correspondiente para el tratamiento de dichas series.

ANÁLISIS ESTADÍSTICO ESTRUCTURAL EMPLEADO

La idea básica de los modelos estructurales de series de tiempo es que ellos pueden ser puestos como modelos de regresión en donde las variables explicativas son funciones del tiempo, con coeficientes que pueden cambiar a través del tiempo. La estimación actual de los coeficientes ó filtrada, se logra poniendo al modelo en forma de espacio de estado y aplicándole luego el denominado Filtro de Kalman (Harvey y Shepard, 1993). Se emplean algoritmos específicos para hacer predicciones y para los suavizados. Esto último significa computar el mejor de los los estimadores en todos los puntos de la muestra usando al conjunto de observaciones. La magnitud por la cual los parámetros pueden variar está gobernada por los llamados hiperparámetros. Estos pueden ser estimados por el método de máxima verosimilitud, construyendo la función específica a optimizar.

El tratamiento estadístico de los modelos estructurales de series de tiempo está basado en la forma de espacio de estado, el filtro de Kalman y el suavizador asociado. La función de verosimilitud se construye a partir del filtro de Kalman en términos de la predicción un paso hacia adelante, y se maximiza con respecto a los hiperparámetros por optimización numérica (Koopman *et al.*, 1995). El vector marcador (“score”) de los parámetros puede obtenerse a través de un algoritmo de suavizado asociado al filtro de Kalman. Una vez que los hiperparámetros fueron estimados, el filtro se usa para conseguir predicciones de los residuos un paso adelante, lo que nos permite calcular los estadísticos para probar normalidad, correlación serial y bondad de ajuste. El suavizador se usa para estimar los componentes que no son observables, como por ejemplo, tendencia y estacionalidad, y para el cálculo de estadísticos que son empleados para detectar observaciones atípicas (“outliers”) y cambios estructurales. El enfoque de espacio de estado es particularmente interesante de ser empleado cuando la serie tiene datos faltantes o han sido incorporados temporalmente. La representación matemática de un modelo de espacio de estado relaciona al vector de disturbios $\{\bullet_t\}$ con el vector de observaciones $\{y_t\}$ a través de un proceso de Markov $\{\bullet_t\}$ y es :

$$\begin{aligned} y_t &= Z_t \bullet_t + G_t \epsilon_t, & \epsilon_t &\sim f(0, H_t) \\ \bullet_t &= T_t \bullet_{t-1} + H_t \eta_t, & \eta_t &\sim f(0, Q_t); & \bullet_0 &\sim f(a_0, P_0), & t=1, \dots, T \end{aligned} \quad (2)$$

donde \bullet_t es el vector de estado de orden $m \times 1$, ϵ_t es un vector de disturbios de orden $k \times 1$ y las matrices del sistema Z_t , T_t , G_t , y H_t tienen dimensiones $N \times m$, $m \times m$, $N \times k$ y $m \times k$ respectivamente. Los disturbios son ruido blanco mutuamente no correlacionados con medio cero y varianza H_t . Cuando se supone normalidad, los disturbios son independientes entre sí. Las matrices G_t y H_t pueden interpretarse como matrices de selección, lo que le brinda generalidad al modelo. Las cuatro matrices son fijas y si en ellas hubiere elementos desconocidos, se incorporan al vector Ψ de hiperparámetros, el que es estimado por máxima verosimilitud. Los estadísticos de prueba usados para la bondad de ajuste son : el BS (de Bowman y Shenton) que emplea estimadores de la asimetría y de la kurtosis de los datos ; el Q (de Box-Ljung) para la autocorrelación seriel y el de DW (de Durbin y Watson), los que junto con otros estadísticos adicionales, califican según sus valores, al modelo como adecuado para ajustar a los datos observados.

Las series originales modeladas en este trabajo constan de $T=405$, $T=535$ y $T=311$ valores promedio mensuales, respectivamente. Gráficamente puede observarse en los tres casos (figuras 1, 3 y 5) presencia de cambios estructurales inducidos por eventuales ciclos e influencia de variables externas (en todos los casos, en el eje de las x se representa el número de mes de observación).

Además se nota un pronunciado comportamiento anómalo en las respectivas funciones CUS, que permiten analizar el comportamiento de la varianza de los datos a través del tiempo (figuras 2, 4 y 6). Si denotamos con s_t , la mencionada función estadística se define clásicamente como sigue :

$$CUS(t) = s_t^{-1} \sum_{j=1}^t (y_j - \bar{y})^2, \text{ donde } j=1, \dots, T, \text{ y el estimador de la varianza muestral es } s_t = [(1/T) \sum_{j=1}^t (y_j - \bar{y})^2]^{1/2} \quad (3)$$

En nuestros casos, la varianza cambia fuertemente en el tiempo, lo cual de no corregirse mediante un modelado apropiado, conduciría a realizar pronósticos muy poco confiables y fuertemente sesgados.

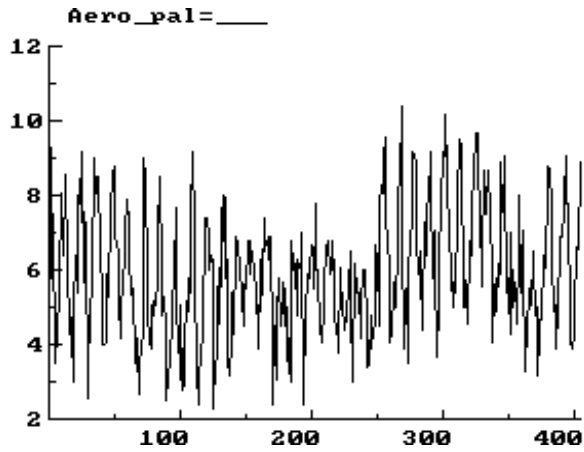


Figura 1 : Serie de heliofania media mensual (hs.) El Palomar

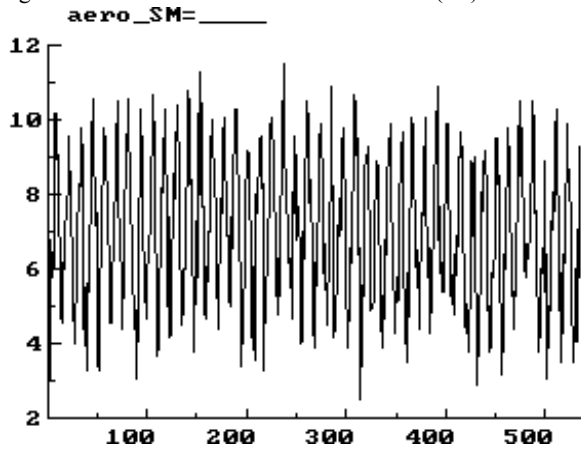


Figura 3 : Serie de heliofania media mensual (hs.) San Miguel

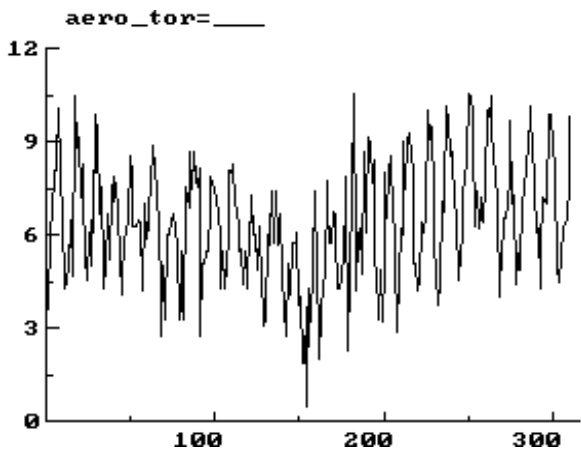


Figura 5 : Serie de heliofania media mensual (hs.) D. Torcuato

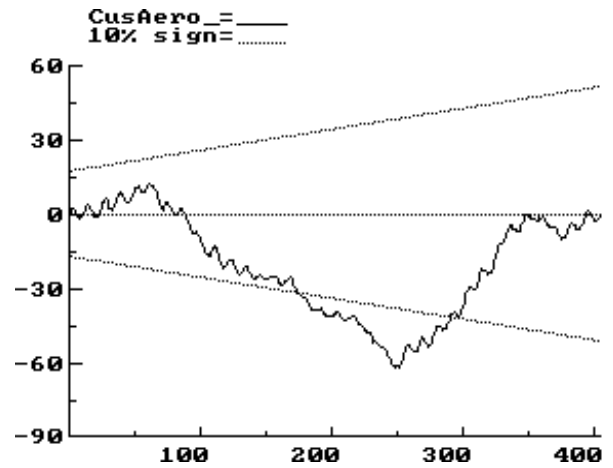


Figura 2 : Función CUS de la serie para El Palomar

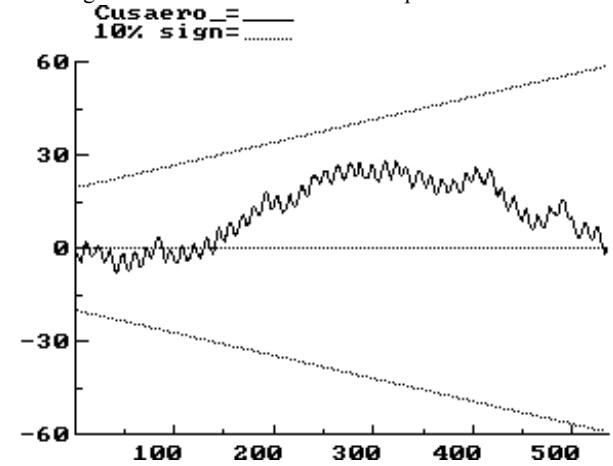


Figura 4 : Función CUS de la serie para San Miguel

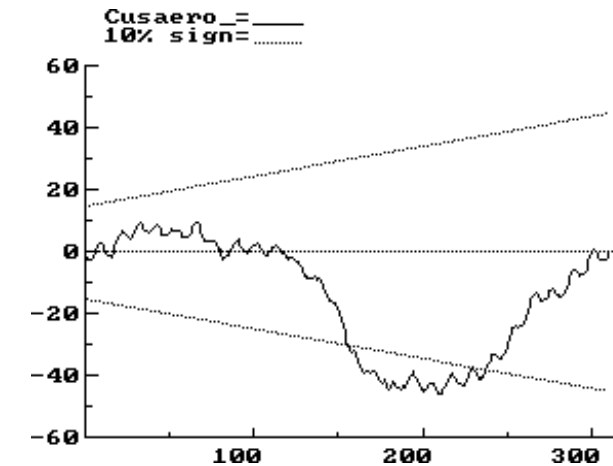


Figura 6 : Función CUS de la serie para D. Torcuato

ANÁLISIS ESTADÍSTICO DE LAS SERIES

Los aspectos observados en las figuras 1 al 6, sugieren trabajar diseñando una metodología diferente a la empleada en el trabajo del año pasado (Avila Blas y Grossi Gallegos , 2002), a fin de maximizar la función de verosimilitud asociada al modelo (2), en el que para estas situaciones, aún usando métodos iterativos apoyados en transformadas rápidas de Fourier, las ecuaciones ortonormales adjuntas son de tipo polinómico de orden alto, y en consecuencia, se incrementa en demasía el número de

hiperparámetros a estimar en los modelos. Es importante comentar aquí, que se ha intentado aplicar metodologías previas a fin de modelizar las series bajo estudio. Las mismas se pueden resumir como sigue :

1) en una primera etapa, hemos empleado el robusto método de **volatilidad estocástica (SV)**, el que puede formularse como sigue :

$$\begin{aligned}
 y_t &= \sigma_t \varepsilon_t = \sigma \varepsilon_t \exp(h_t/2), \quad \bullet_t \sim \text{IID}(0, 1), \quad t=1, \dots, T \\
 h_t &= \phi h_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad |\phi| < 1
 \end{aligned}
 \tag{4}$$

El término σ^2 es un factor de escala, ϕ es un parámetro y η_t es un término de disturbio que se considera no correlacionado con ε_t . Pero los estadísticos de la bondad de ajuste correspondientes, no resultaron satisfactorios.

- 2) tampoco pudo tratarse efectivamente a las series, tomando a la serie de disturbios como combinaciones lineales de distribuciones normales truncadas, ni tampoco de tipo t de Student.
- 3) Sin embargo, luego de varios intentos fallidos, procedimos a tomar nuevamente el antes mencionado “índice de contribución relativa”, pero esta vez, previamente se diseñó para cada caso, una nueva serie y^* , definiendo un tipo nuevo de volatilidad estocástica, a la que denominamos **volatilidad estocástica de segundo orden**, y que puede formularse de la siguiente manera :

$$\begin{aligned}
 y_t^* &= \sigma_t \varepsilon_t = \sigma \varepsilon_t \exp(h_t/2) + g(\varepsilon_t); \text{ donde } g \text{ es un polinomio de grado 2, y } \bullet_t \sim \text{IID}(0, 1), \quad t=1, \dots, T \\
 h_t &= \phi h_{t-1} + g(\varepsilon_{t-1}) + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad |\phi| < 1
 \end{aligned}
 \tag{5}$$

la segunda ecuación define ahora una **caminata aleatoria de orden 2 implícita**, por la intervención de la función g . La componente de suavizado h_t mediante un sencillo método analítico puede ahora estimarse en a lo sumo dos etapas : una de aproximación en el sentido de la optimización de una función con ligas y otra mediante una acotación con ponderaciones dependientes del máximo nivel de variabilidad observado en el período de estudio de cada serie. La función de densidad g tiene como objetivo en este caso, actuar como una función de control, en el sentido que delimita las posibles variaciones aleatorias de la función h .

Una vez conseguida la nueva serie de disturbios (la asociada a la serie y^*), y expresando $\varepsilon_t = y_t - Z_t \bullet_t$, podemos aplicar el índice de contribución relativa de la densidad g a la distribución conjunta del vector de estado α , y el vector de datos valuado en el tiempo máximo y_T , obteniendo la siguiente expresión :

$$I = \left\{ \frac{\partial \log g(y_t - Z_t \bullet_t)}{\partial \bullet_t} \right\} \left\{ \frac{\partial P(y_T, \bullet_t)}{\partial \bullet_t} \right\} = Z_t^* (y_t - Z_t^* \bullet_t) g^2(y_t - Z_t^* \bullet_t) g^2(y_t - Z_t \bullet_t)
 \tag{6}$$

donde el factor $g^2(y_t - Z_t^* \bullet_t)$ puede acotarse para cada valor experimental $\varepsilon_t^* = y_t - Z_t \bullet_t^*$

Como se desconoce la varianza residual σ_ε^2 , proponemos como un estimador con buenas propiedades (asintóticamente insesgado, convergente en probabilidad y de máxima eficiencia) a $S_\varepsilon^2(\varepsilon_t)$, que cumple con la condición :

$$E\left\{ \frac{(T+1)}{T} S_\varepsilon^2(\varepsilon_t) \right\} = E\left\{ \frac{(T+1)}{T} g^2(y_t - Z_t^* \bullet_t) \right\} \rightarrow \sigma_\varepsilon^2, \quad \text{cuando } T \rightarrow \infty
 \tag{7}$$

Podemos en consecuencia expresar el índice de contribución relativa como :

$$I = \left[\frac{(T+1)}{T} S_\varepsilon^2(\varepsilon_t) \right] (y_t - Z_t \bullet_t^*) Z_t^*
 \tag{8}$$

el que tiene ahora una distribución asintótica Gaussiana estandarizada, y por lo tanto se le puede aplicar el método tradicional de suavizado y filtrado de Kalman (SFK) (a la serie transformada por el modelo de volatilidad estocástica de orden 2), para realizar las estimaciones de los hiperparámetros asociados al modelo original en cada serie. El nivel de confiabilidad γ adoptado para el tratamiento de los datos fué del 95%.

Aplicando la transformación dada por (8) las series modificadas y empleando el soft STAMP 5.0 (Koopman *et al.*, 1995), obtenemos las estimaciones de los hiperparámetros $(\sigma, \phi, \sigma_\eta^2)$, como así también las cotas $C(g)$ para la función de control g , y el número de períodos de predicción que el respectivo modelo de cada serie (N), que se muestran en la Tabla 1.

Localidad	σ	ϕ	σ_n^2	C(g)	N	γ
El Palomar	1.186	0.991	0.075	0.019	39	95%
San Miguel	1.424	0.997	0.056	0.023	45	95%
Don Torcuato	1.456	0.971	0.034	0.013	19	95%

Tabla 1 : Valores característicos estimados de los hiperparámetros y cotas de la función de control para el modelado realizado.

Es importante destacar que para cada serie bajo estudio, la convergencia de los métodos iterativos de estimación empleados, se logró usando un promedio de 17 iteraciones, obteniendo el máximo de la función de verosimilitud asociada al proceso y el cambio relativo para la estimación de los hiperparámetros y el vector de estado, con un error medio del orden de 10^{-16}

Si tenemos en cuenta que una de las etapas más difíciles del procedimiento en la bondad de ajuste de este tipo de modelos, es la estimación de los hiperparámetros, como así también las cotas de la función de ajuste g, los resultados antes mencionados son muy satisfactorios y presentan robustez apropiada, que los hace muy confiables.

Los resultados analíticos destacados anteriormente pueden ser corroborados con las gráficas asociadas a la predicción (figuras 7, 9 y 11) y sus respectivas funciones CUS (figuras 8, 10 y 12), que muestran un claro control en la variabilidad de errores de predicción, al quedar cada gráfica completamente contenida en las bandas de confiabilidad.

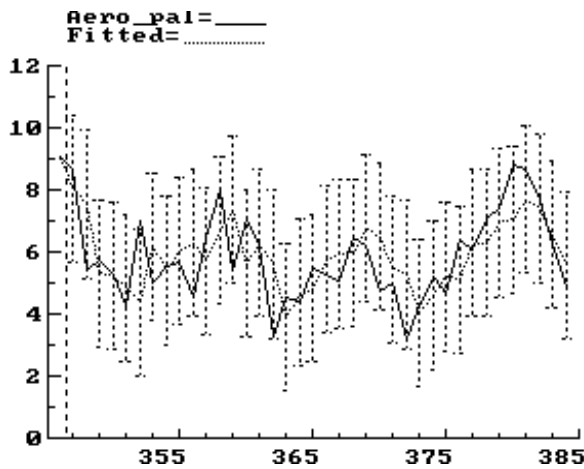


Figura 7 : predicciones 39 meses hacia adelante : El Palomar Palomar

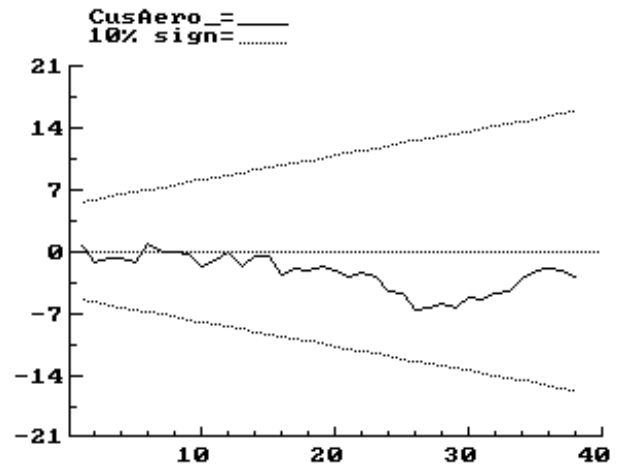


Figura 8 : Función CUS para el modelo estructural : El

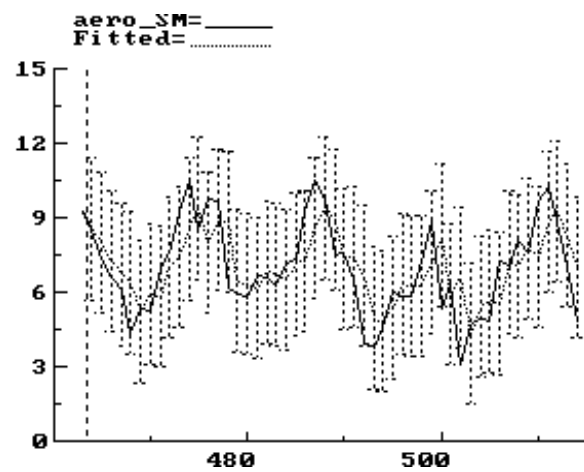


Figura 9 : predicciones 39 meses hacia adelante : San Miguel

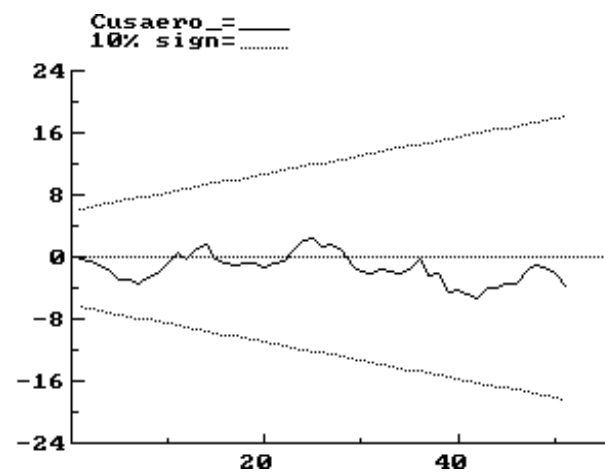


Figura 10 : Función CUS para el modelo estructural: San Miguel

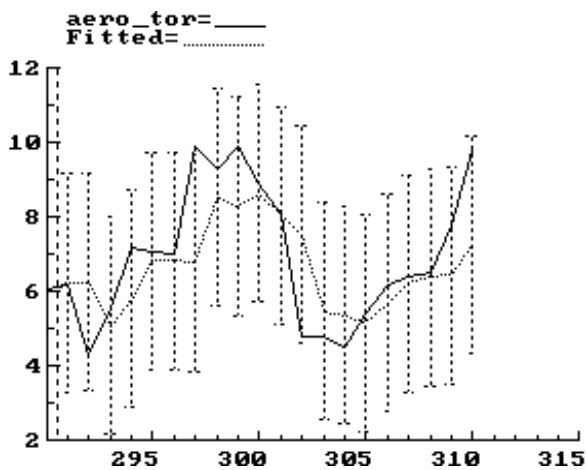


Figura 11: predicciones 19 meses hacia adelante: D. Torcuato Torcuato

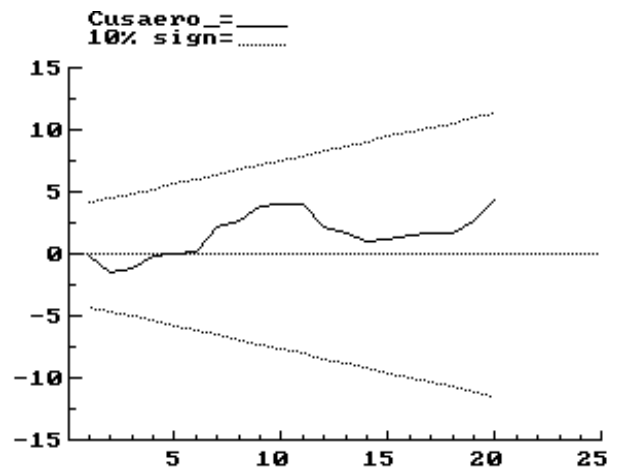


Figura 12 : Función CUS para el modelo estructural: D.

CONCLUSIÓN

Mediante el empleo del tratamiento estructural de series de tiempo, se consiguió modelar tres series de valores de promedios mensuales de heliofanía medidas en estaciones del S.M.N, con una bondad de ajuste muy buena (95%). Esto permite realizar estimaciones a largo y mediano plazo con una confiabilidad alta. Este tipo de procedimiento difiere de los que se encuentran tratados o mencionados en la bibliografía disponible, y tienen vasta difusión entre los investigadores. Estos nuevos resultados obtenidos se agregan a los conseguidos en años anteriores, para diferentes localidades del país, y amplían el panorama del comportamiento de la variable bajo estudio, por lo que podemos pensar en continuar recomendando su uso de una manera general para la simulación y generación de series de valores sintéticos a los fines de una simulación detallada del comportamiento de aquellos sistemas de aprovechamiento de la energía solar que así lo requieran.

REFERENCIAS

- Avila Blas O.J. y Grossi Gallegos H. (2002). Modelos estadísticos estructurales de series de irradiación solar global diaria para Córdoba, Marcos Juárez y Paraná. *Avances en Energías Renovables y Medio Ambiente*, **6**, 2, 11.07-11.11
- Harvey A. C. y Shepard N. (1993). Estructural Time Series Models. En *Handbook of Statistics*, Elsevier Science Publishers B.V (Editores), **11**, pp. 261-302.
- Koopman S. J., Harvey A. C., Doornik J. A. y Shepard N. (1995). *STAMP 5.0, Structural Time Series Analyser, Modeller and Predictor*. 1a. edición. Chapman and Hall, London.
- Suehrcke H. (2000) On the relationship between duration of sunshine and solar radiation on the Earth's surface: Ångström's equation revisited. *Solar Energy* **68**, 5, 417-425

ABSTRACT- In this paper a statistical study for the treatment of three monthly average series of sunshine hours, by using structural analysis method is done, with data corresponding to El Palomar, San Miguel and Don Torcuato locations. In order to present models for generating synthetic monthly average values of for these locations, the statistical theory was developed which allow us to make forecasting with a high confidence level (95%)

Keywords : sunshine hour, monthly averages, statistic, structural models, forecasting