

## MODELOS ESTADÍSTICOS ESTRUCTURALES DE SERIES DE IRRADIACIÓN SOLAR GLOBAL DIARIA PARA CÓRDOBA, MARCOS JUÁREZ Y PARANÁ

Orlando José Avila Blas<sup>1</sup> y Hugo Grossi Gallegos<sup>2</sup>

<sup>1</sup>Av. Bolivia 5150, (4400) Salta, Argentina, Tel. : (0387) 4255385, Email : [oblas@unsa.edu.ar](mailto:oblas@unsa.edu.ar)

<sup>2</sup>Rutas 5 y 7, (6700) Luján, Buenos Aires, Argentina, Fax: (2323) 425795, Email: [grossi@mail.unlu.edu.ar](mailto:grossi@mail.unlu.edu.ar)

**RESUMEN** - En el presente trabajo se realiza un estudio estadístico de series de irradiación solar global diaria, empleando la metodología del análisis estructural, para datos tomados en las ciudades de Córdoba, Marcos Juárez y Paraná. Se ha desarrollado la teoría estadística correspondiente a fin de presentar respectivos modelos para poder generar sintéticamente valores de radiación diarios, lo que permite realizar pronósticos con una muy alta confiabilidad (99%).

**Palabras clave:** irradiación solar global, estadística, modelado estructural, pronósticos.

### INTRODUCCION

Para las predicciones a largo plazo del comportamiento de los sistemas de aprovechamiento de la energía solar a menudo sólo hace falta un conocimiento de los valores medios mensuales de la irradiación global diaria disponible en una localidad (por ejemplo, sistemas domésticos de colectores planos para calentamiento de agua). Pero para otras aplicaciones es necesario una simulación detallada para la que se requieren datos diarios (o aún horarios) y este tipo de información generalmente no se dispone o las series existentes tienen poco desarrollo temporal. Esto dio lugar a que se buscaran alternativas que permitieran satisfacer esta demanda.

Pueden diferenciarse tres etapas en el siglo pasado dentro de la estadística de la radiación solar. La primera de ellas, en la década del '20, fue la marcada por la contribución de Ångström (1924) quien vinculó la información de horas de insolación con la radiación global; la segunda, que dio un verdadero impulso a esta rama de la investigación, fue concretada por los pioneros Liu y Jordan (1962), quienes analizaron la distribución de probabilidad de la radiación solar recibida en varias localidades canadienses y norteamericanas descubriendo una serie de regularidades en los datos de irradiación global y difusa, tanto diarios como horarios. Finalmente, a fines de los años '70 y comienzos de los '80, los investigadores comenzaron a utilizar la teoría de *modelización estocástica* para generar series temporales (series sintéticas, indiferenciables de las reales) de valores diarios de irradiación (Brinkworth, 1979; Aguiar *et al.*, 1988; Graham *et al.*, 1988).

En los años más recientes, se han propuesto métodos estadísticos de generación de series de irradiación horaria (Aguiar, 1992), utilizando modelos basados en el tratamiento de Box y Jenkins, es decir, modelos ARIMA(p,d,q), también se han modelado series para la ciudad de Salta, específicamente para la localidad de Cerrillos, haciendo uso de la metodología de los modelos estadísticos estructurales (Avila Blas *et al.*, 1999, 2000). En este último caso, se obtuvieron muy buenos resultados, empleando el denominado modelo de "volatilidad estocástica", permitiendo que la varianza asociada a los disturbios de la serie pueda variar aleatoriamente y sin la necesidad de imponer sobre ellos la condición de tener una distribución normal o Guassiana (hecho muy relevante, ya que los datos asociados a radiación no poseen este tipo de distribución). En el presente trabajo se han modelado series de irradiación global diaria para las ciudades de Córdoba, Marcos Juárez y Paraná a partir de datos medidos por la Red Solarimétrica. Cada una de ellas tiene una longitud distinta y además características especiales que las hacen diferentes de la serie de valores estudiada para Salta, aunque finalmente, salvo el enfoque especial introducido para el tratamiento de ellas mediante el suavizado y filtrado de Kalman, sus respectivos modelos pertenecen a la "familia de volatilidad estocástica".

### ANÁLISIS ESTADÍSTICO ESTRUCTURAL EMPLEADO

La idea básica de los modelos estructurales de series temporales 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. mejor de 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; éstos 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. 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

<sup>1</sup> Probabilidades y Estadística, Departamento de Matemática, Fac. Cs. Exactas - UNSa

<sup>2</sup> División Física, Universidad Nacional de Luján, Miembro de la Carrera del Investigador del CONICET

cuando la serie tiene datos faltantes o han sido incorporados temporalmente. En este trabajo, se ha tomado como representación matemática del **modelo general de espacio de estado** relaciona al vector de disturbios  $\{\varepsilon_t\}$  con el vector de observaciones  $\{y_t\}$  a través de un proceso de Markov  $\{\alpha_t\}$ , a :

$$\begin{aligned} y_t &= Z_t \alpha_t + G_t \varepsilon_t, & \varepsilon_t &\sim f(0, H_t) \\ \alpha_t &= T_t \alpha_{t-1} + H_t \eta_t, & \eta_t &\sim f(0, Q_t); \quad \alpha_0 \sim f(a_0, P_0), \quad t=1, \dots, T \end{aligned} \quad (1)$$

donde  $\alpha_t$  es el vector de estado de orden  $m \times 1$ ,  $\varepsilon_t$  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. La densidad  $f$  no es necesariamente normal o Gaussiana. 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  $\Psi$  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 serial 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 analizadas en este trabajo, constan de  $T=5053$ ,  $T=2699$  y  $T=8402$  valores promedios diarios de irradiación solar global diaria, respectivamente. La falta de normalidad en los datos, la presencia de outliers (en poca cantidad y de valores grandes) y períodos sin mediciones, que es el caso de las series aquí tratadas (Figuras 1, 3 y 5) (en el eje de las  $x$ , se representa el número de día de observación). Se observa además que los respectivos correlogramas (Figuras 2, 4 y 6) muestran un proceso de convergencia excesivamente lento (en estos gráficos, el eje de las  $x$  es adimensional, por tratarse de un índice relativo).

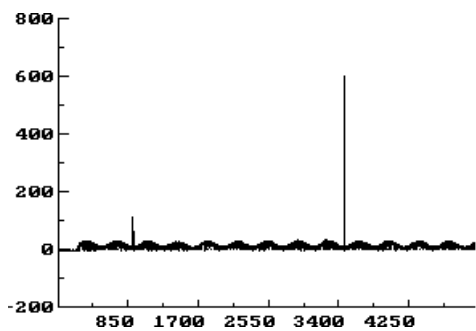


Figura 1: Serie de irradiación para Córdoba (MJ/m<sup>2</sup>)

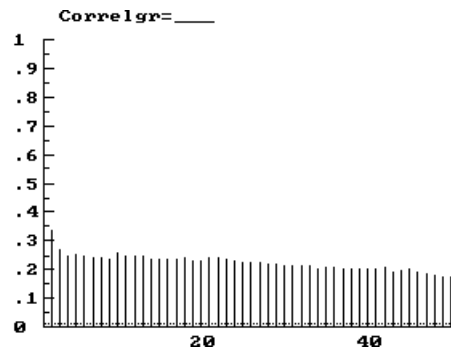


Figura 2: Correlograma de la serie para Córdoba

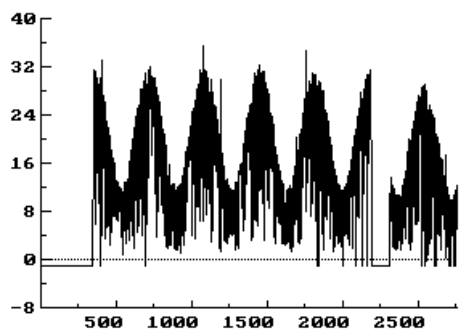


Figura 3: Serie de irradiación para Marcos Juárez (MJ/m<sup>2</sup>)

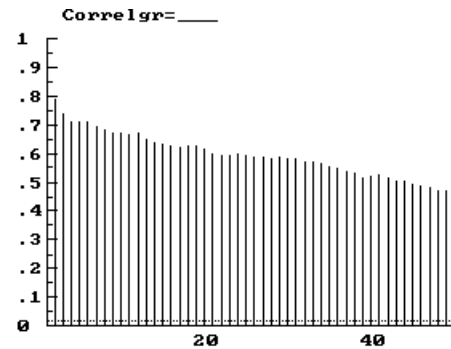


Figura 4: Correlograma de la serie para Marcos Juárez

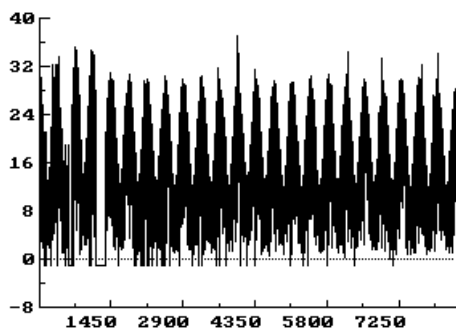


Figura 5: Serie de irradiación para Paraná (MJ/m<sup>2</sup>)

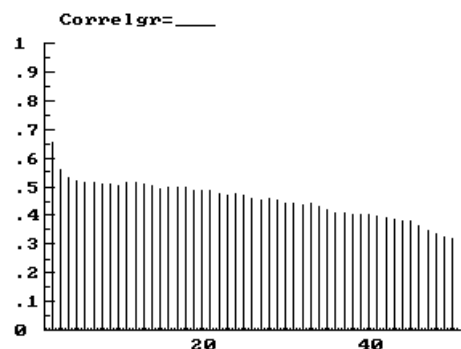


Figura 6: Correlograma de la serie para Paraná

Estos aspectos sugieren el empleo de un tratamiento más complejo que implica el uso de una función de verosimilitud, a maximizar por métodos iterativos, dado que las ecuaciones ortogonales asociadas no son lineales. En este sentido, hemos intentado varios modelos previos para cada serie, entre los cuales se pueden citar a:

- 1) el uso de modelos de volatilidad estocástica simple, aplicados en forma directa a la serie original:

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

donde el término  $\sigma^2$  es un factor de escala,  $\phi$  es un parámetro y  $\eta_t$  es un término de disturbio que se considera no correlacionado con  $\varepsilon_t$ . Los valores  $\phi$ ,  $\sigma_\eta^2$  y  $\sigma$ , son en este modelo los hiperparámetros a estimar.

- 2) el tratamiento de las series de disturbios mediante procesos iterativos, donde se supone que éstos tienen una densidad de probabilidad que es una combinación lineal de dos distribuciones normales ó de una distribución normal y una distribución t de Student.

Lamentablemente, ninguno de éstos modelos dieron resultados estadísticos confiables, especialmente al analizar la información brindada por los estadísticos usados para mediar la bondad del ajuste (BS, Q y DW).

Un tercer tratamiento, el cual resultó ser la solución a nuestro problema de modelado, consiste en tratar a las series de disturbios como una combinación lineal de dos distribuciones de tipo t de Student, de la forma

$$h(\varepsilon_t) = (1-\beta) h_1(t) + \beta h_2(t) \quad (3)$$

donde las funciones  $h_1$  y  $h_2$  son densidades de Student con  $n_1$  y  $n_2$  grados de libertad respectivamente, es decir:

$$h_i(\varepsilon_t) = [c(n_i) / \sigma_\varepsilon] [1 + \varepsilon_t^2 (c(n_i) - 2)^{-1} \sigma_\varepsilon^{-2}]^{-c(n_i)/2}, \quad n_i > 2, \quad i=1,2 \quad (4)$$

en esta expresión, se considera que las  $c(n_i)$  son constantes y se escriben de este modo, de manera que la varianza residual  $\sigma_\varepsilon^2$  se mantiene también constante, a fin de independizar relativamente la estimación de estos parámetros en el proceso de estimación de los hiperparámetros asociados al modelo estructural correspondiente.

La densidad  $h_1$  tiene como objetivo principal representar la mayor parte de la distribución real de los disturbios y  $h_2$  tiene como función manejar solamente un número relativamente pequeño de datos atípicos o "outliers" numéricamente grandes.

Teniendo en cuenta que podemos expresar a los disturbios como  $\varepsilon_t = y_t - Z_t \alpha_t$ , es importante determinar el índice de contribución relativa de la densidad compuesta  $h$  a la distribución conjunta del vector de estado  $\alpha$  y el vector de datos  $y_T$ , el cual se determina como

$$I = \{\partial \log h(y_t - Z_t \alpha_t) / \partial \alpha_t\} / \{\partial P(\alpha_t, y_T) / \partial \alpha_t\} = Z_t' h(\varepsilon_t)^{-1} (y_t - Z_t \alpha_t) [(1-\beta) c(n_1) h_1(t) + \beta c(n_2) h_2(t)] \quad (5)$$

Y entonces, proponiendo la ecuación del estimador del desvío estándar (S) de los disturbios como

$$[S_\varepsilon^2(\varepsilon_t)]^{-1} = (n_1 + n_2 + 2) \sigma_\varepsilon^2 / (n_1 + n_2 - 4) \quad (6)$$

y tomando el valor  $\varepsilon_t^* = y_t - Z_t \alpha_t^*$ , a partir del valor experimental  $\alpha_t^*$ , la ecuación de contribución relativa adopta la forma

$$I = [S_\varepsilon^2(\varepsilon_t)]^{-1} Z_t' (y_t - Z_t \alpha_t) \quad (7)$$

la que ahora tiene una forma cuasi-Gaussiana estándar, y en consecuencia podemos aplicarle el método de suavizado y filtrado de Kalman (SFK) a la serie transformada por el modelo de volatilidad estocástica, para realizar las estimaciones asociadas al modelo original. En el caso especial de las series bajo estudio, se ha logrado trabajar cómodamente con un muy alto nivel de confiabilidad (99%), es decir, un nivel de riesgo del 1%.

## ANÁLISIS ESTADÍSTICO DE LAS SERIES

Una vez realizado el tratamiento estadístico descrito en las ecuaciones (5) a (7), para cada serie de datos, hemos llegado a los correspondientes modelos finales, que presentan la siguiente estructura genérica (lo que queda corroborado por los valores de los estadísticos asociados a la bondad de ajuste : BS, Q y DW).

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t = \sigma \varepsilon_t \exp(h_t/2), \quad \varepsilon_t \sim \text{IID}(0, 1), \quad t=1, \dots, T \\ h_t &= \phi h_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad |\phi| \leq 1 \end{aligned} \quad (8)$$

El término  $\sigma^2$  es un factor de escala,  $\phi$  es un parámetro y  $\eta_t$  es un término de disturbio que se considera no correlacionado con  $\varepsilon_t$ .

En este tipo de modelo no podemos realizar directamente la maximización de la función de verosimilitud a fin de aplicar el Filtro de Kalman como en los modelos básicos. Sin embargo, podemos aplicar un método de cuasi-máxima verosimilitud (QML), fácil de aplicar y razonablemente eficiente. El mismo está basado en la siguiente aproximación en serie de Taylor :

$$\log y_t^2 \cong \log (y_t^2 + 0.02 s_y^2) - 0.02 s_y^2 / (y_t^2 + 0.02 s_y^2), \quad t=1, \dots, T \quad (9)$$

donde  $s_y^2$  es la varianza de los datos.

Finalmente, los modelos particulares para las 3 series bajo estudio, y que responden a la ecuación (8), tienen respectivamente los valores característicos que se indican en la Tabla 1.

Localidad	$\sigma$	$\phi$	$\sigma_\eta^2$	Cantidad de períodos para pronósticos	Nivel de significancia
Córdoba	1.324	0.992	0.780	53	1%
Marcos Juárez	1.444	0.997	0.173	60	1%
Paraná	1.301	0.986	0.107	50	1%

Tabla 1: Valores característicos estimados de los hiperparámetros para el modelo estructural final.

Cabe destacar que para cada una de las tres series se logró en promedio obtener la convergencia apropiada, al maximizar la función de verosimilitud, en un promedio de 30 iteraciones. Además, el cambio relativo promedio para todos los hiperparámetros fue del orden de  $10^{-16}$ , y para el vector de estado, del orden de  $10^{-10}$

Teniendo en cuenta que la etapa más dificultosa del procedimiento de ajuste y estimación en los modelos estructurales de series de tiempo es usualmente la estimación de los hiperparámetros, estos resultados son altamente satisfactorios

Los resultados analíticos se pueden corroborar con las gráficas para el tratamiento residual (Figuras 7, 9 y 11) y de pronóstico ( comparadas con las series originales, en Figuras 8, 10 y 12), una vez que se les aplicó a las series originales el tratamiento estadístico anteriormente indicado.

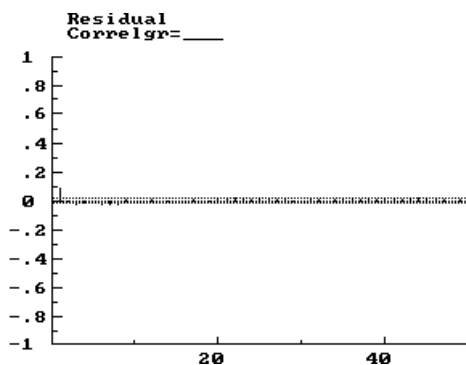


Figura 7: Correlograma residual para Córdoba

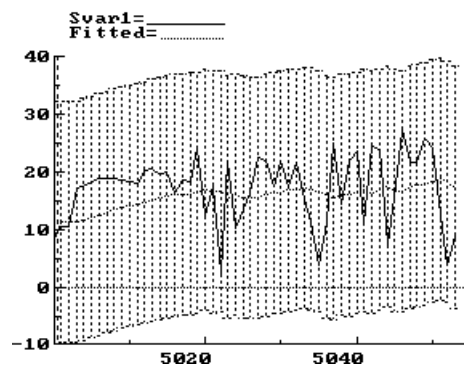


Figura 8: Predicciones 53 días hacia delante (Córdoba)

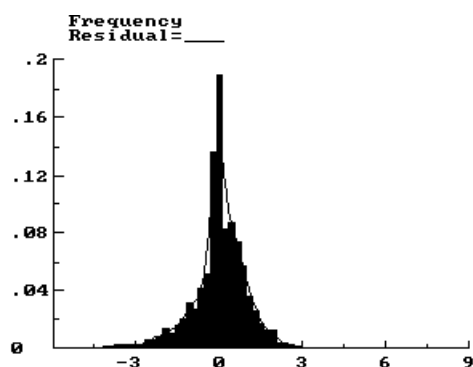


Figura 9: Distribución empírica residual para Marcos Juárez

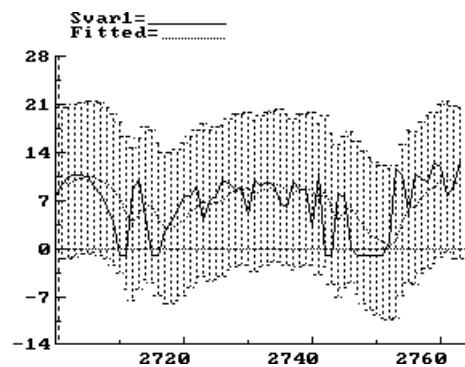


Figura 10: Predicciones 60 días hacia delante (M. Juárez)

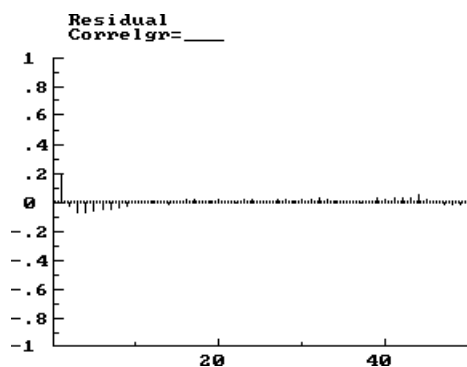


Figura 11 : Correlograma residual para Paraná

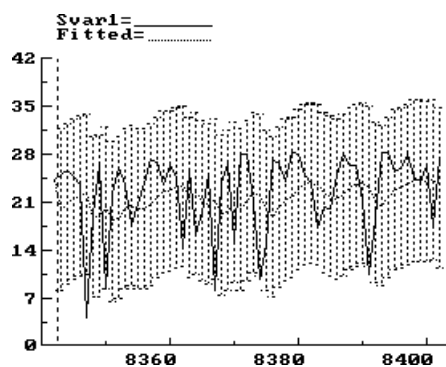


Figura 12: Predicciones 50 días hacia delante (Paraná)

## CONCLUSIÓN

Mediante el empleo del tratamiento estructural de series de tiempo, se consiguió modelar series de valores de irradiación global diaria medidas por la Red Solarimétrica, con una bondad de ajuste muy buena, lo permite realizar estimaciones a corto plazo con una confiabilidad alta (99%), además de poder realizar estas predicciones a largo plazo (todas superiores a los 50 días). El procedimiento utilizado difiere de los que habitualmente se encuentra en la bibliografía citada y que tienen vasta difusión entre los investigadores; además, el haberlo ensayado ya con muy buenos resultados en cuatro ubicaciones que presentan diferencias apreciables por la longitud de sus series y la ubicación geográfica de las estaciones de medición permite pensar en la recomendación del uso de esta metodología de manera más general para la simulación detallada del comportamiento de sistemas de aprovechamiento de la energía solar que así lo requieren.

## REFERENCIAS

- Abril J. C. (2000) *Análisis Estadístico de Series de Tiempo Basado en Modelos de Espacio de Estado*, 1ª. Edición. EUDEBA, Buenos Aires.
- Aguiar R. y Collares Pereira M. (1992) Tag: A time dependent, autoregressive, gaussian model for generating synthetic hourly radiation. *Solar Energy* **49**, 3, 167-174.
- Aguiar, R.J., Collares-Pereira, M. and Conde, J.P. (1988) Simple procedure for generating sequences of daily radiation values using a library of Markov transitions matrices. *Solar Energy* vol. 40 No. 3, pp. 269-279.
- Ångström, A. (1924) Solar and terrestrial radiation. *Q. J. R. Meteorol. Soc.* vol. 50, pp. 121-125.
- Avila Blas O. J., Abril J. C. y Lesino G. (1999) Análisis estadístico estructural de series de radiación diarias. *Avances en Energías Renovables y Medio Ambiente* **3**, 2, 11.17-11.20.
- Avila Blas O. J., Abril J. C. y Lesino G. (2000) Radiación y temperaturas diarias: Un modelo de correlación estructural. *Avances en Energías Renovables y Medio Ambiente* **4**, 2, 11.31-11.36
- Brinkworth, B.J. (1979) Autocorrelation and stochastic modelling of insolation sequences. *Solar Energy* vol. 19, pp. 343-348.
- Graham, V. A., Hollands, K.G.T. and Unny, T.E. (1988) A time series model for Kt with application to global synthetic weather generation. *Solar Energy* vol. 40 No. 2, pp. 83-92.
- Harvey A. y Shepard N. (1993) Structural Time Series Models, *Handbook of Statistics* **11**, 621-302.
- Liu, B.Y.H. and Jordan, R.C. (1960) The interrelationship and characteristic distribution of direct, diffuse and total radiation. *Solar Energy* vol. 4 No. 3, pp. 1-19.

**ABSTRACT** – In this paper a statistical time series treatment for daily global solar irradiance by using structural analysis method is done, with data corresponding to Córdoba, Marcos Juárez and Paraná cities. In order to present models for generating synthetic daily global irradiance values for these cities, the statistical theory was developed which allow us to make forecasting with a high confidence level (99%).