

RADIACION Y TEMPERATURA DIARIAS: UN MODELO DE CORRELACION ESTRUCTURAL

O. J. Avila Blas¹, J. C. Abril², G. Lesino³

Av. Bolivia 5150 - 4400 Salta

Tel. : (0387) 4255385, (0381) 4364093

Email : oblas@unsa.edu.ar, jabril@herrera.unt.edu.ar, lesino@unsa.edu.ar

RESUMEN

Continuando una línea de investigación iniciada en 1998, en el presente trabajo se realiza un estudio estadístico bivariado de series de radiación promedio y temperatura promedio diarias empleando la metodología del análisis estructural, introducidos por estadísticos contemporáneos. Se presenta un modelo de correlación entre ambas series para poder generar sintéticamente valores de radiación diarios a partir de valores de temperaturas diarias, efectuando pronósticos con una alta confiabilidad (95%).

Palabras clave : Radiación, Temperatura, Correlación, Análisis Estructural.

INTRODUCCION.

Para el diseño de sistemas solares, se han desarrollado programas computacionales de simulación de diferentes niveles de sofisticación. Existen programas aproximados, que en realidad son herramientas de prediseño, mediante los cuales se predice la producción y la eficiencia con un error entre el 10 y el 20%. La mejora en el ajuste de los programas más precisos proviene de una mejor descripción física de los procesos involucrados, lo que obliga a mejorar también la calidad de los datos meteorológicos con que dichos programas se ejecutan. Dado que son cálculos predictivos, es obvio que no se cuenta con el dato climático real.

Para la simulación de los sistemas solares son necesarias varias variables meteorológicas, tales como radiación solar, temperatura ambiente y vientos, muchas veces en promedios horarios. Estas variables están correlacionadas entre sí y en el tiempo.

En los últimos años se han propuesto métodos estadísticos de generación de series de algunas variables de interés : temperatura de superficie (Avila Blas, 1997), radiación horaria (Aguar y Collares Pereira, 1992), utilizando modelos basados en el tratamiento de Box y Jenkins, es decir, modelos ARIMA(p,d,q) y también modelos estructurales de radiación diaria (Avila Blas *et al*, 1999).

No se conocen hasta el momento trabajos de modelado multivariado de series de tiempo con la metodología del enfoque estructural introducido por estadísticos contemporáneos (Harvey y Shepard, 1993 ; Abril, 1997). La idea central de este trabajo en iniciar este tipo de estudio, al caso de una serie de valores de radiación promedio diaria y otra de valores de temperatura promedio diaria, las que corresponden a mediciones realizadas entre los años 1981 y 1996 en Salta Capital. Los datos fueron brindados por el Servicio Meteorológico Nacional con sede en el Aeropuerto Salta.

ANALISIS ESTADISTICO ESTRUCTURAL BIVARIADO

La representación matemática de las ecuaciones que describen un modelo estructural multivariado, tienen un esquema similar al de una ecuación unidimensional, con la particularidad de que el vector de observaciones $\{y_t\}$ tiene dimensión $T \times 1$, donde T es la cantidad de observaciones de la serie a tratar. La representación matemática de un modelo 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 $\{\alpha_t\}$ y tiene la forma

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

donde α_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.

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

² INIE, Instituto UNT - CONICET

³ INENCO, Instituto UNSa - CONICET

Recordemos que 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. 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.

Cuando se trata de datos diarios, los modelos estructurales pueden ser extendidos de tal forma de permitir el manejo de los mismos introduciendo un componente diario. Este se modela en la misma forma que un componente estacional y puede permitírsele que varíe en el tiempo. Otros componentes como la tendencia y la estacionalidad anual pueden ser introducidos como lo explicado en el modelo general. La falta de normalidad en los datos, que es el caso de nuestro serie, requiere de un tratamiento más particular que implica el uso de una función de verosimilitud que se maximiza por métodos iterativos, dado que las ecuaciones normales asociadas no son lineales. En este sentido se puede intentar tratar a los datos como provenientes de una familia exponencial ; o bien, cuando los datos tienen una distribución de “colas pesadas”, se puede emplear combinaciones de distribuciones normales con la t de Student.

Las series originales tratadas aquí constan de N=5842 valores diarios cada una. Pero, dada la ventaja del tratamiento estructural, podemos trabajar con una serie sustancialmente más pequeña, por ejemplo tomando el par de años 1982 y 1983 ; en particular los T=360 primeros valores, a fin de poder compararlos con los datos predichos por la serie modelada. En estos años se observa la presencia de datos “outliers” o extremos, los que fueron tratados con la incorporación de variable explicativas, previo al proceso de análisis de correlación estructural.

El modelo estructural para la serie de radiación diaria (Rad) obtenida anteriormente (Avila Blas *et al.*, 1999) corresponde al de un **modelo de varianza o volatilidad estocástica (SV)**, cuya estructura viene dada por:

$$y_{1t} = \sigma_{1t} \varepsilon_{1t} = 3.43 \varepsilon_{1t} \exp(h_{1t}/2), \quad \varepsilon_{1t} \sim \text{IID}(0, 1), \quad t=1, \dots, 730; \quad \text{donde } h_{1t} = 0.955 h_{1t-1} + \eta_{1t}, \quad \eta_{1t} \sim \text{NID}(0, 0.05) \quad (2)$$

El término $\sigma = 3.43$ es un factor de escala, $\phi = 0.995$ es un parámetro y η_t es un término de disturbio que se considera no correlacionado con ε_t .

En una primera etapa del presente trabajo se consiguió un modelo análogo al establecido por la ecuación (2) para ajustar los datos de la serie correspondiente a los valores de temperatura promedio diaria (Tprom). El mismo responde a la ecuación :

$$y_{2t} = \sigma_{2t} \varepsilon_{2t} = 3.82 \varepsilon_{2t} \exp(h_{2t}/2), \quad \varepsilon_{2t} \sim \text{IID}(0, 1), \quad t=1, \dots, 730; \quad \text{donde } h_{2t} = 0.974 h_{2t-1} + \eta_{2t}, \quad \eta_{2t} \sim \text{NID}(0, 0.04) \quad (3)$$

Nótese que las distribuciones empíricas de las variables de estudio, no son gaussianas y presentan una fuerte asimetría por la izquierda. En cada gráfico, cuando no se indiquen unidades, es porque la variable es adimensional.

Las similitudes así halladas pueden corroborarse observando las representaciones gráficas de las series a las que se le aplicó la transformación de volatilidad (SV), en las figuras 1 y 3, con sus respectivas densidades en las figuras 2 y 4

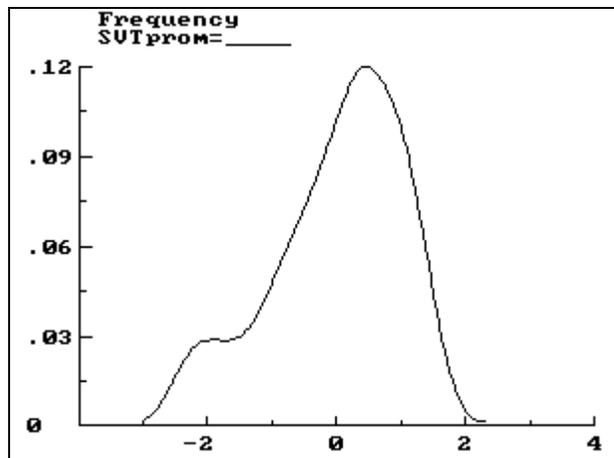
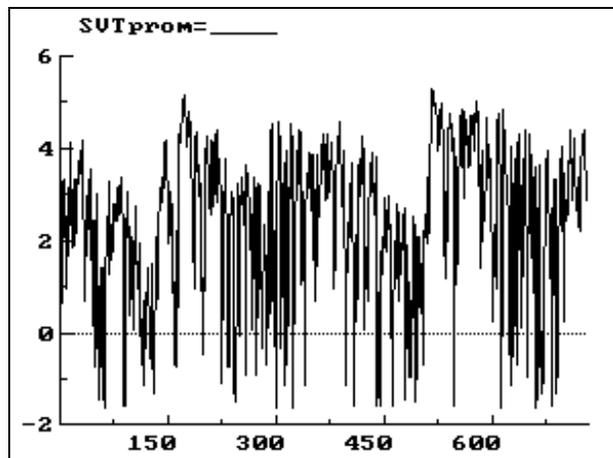


Figura 1: Serie SV de temperatura diaria (°C), (eje x en días)

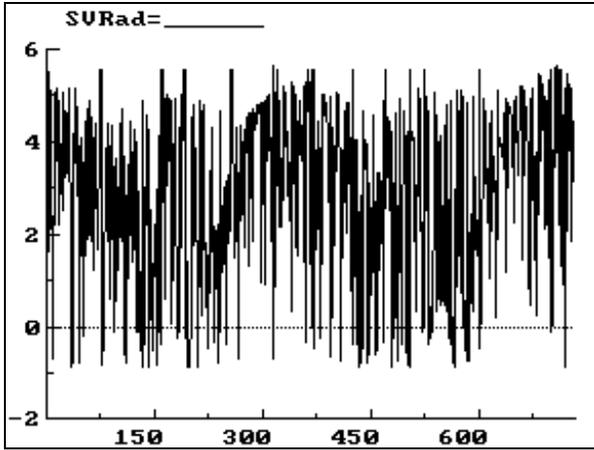


Figura 2 : densidad SV de temperatura diaria, (eje x en °C)

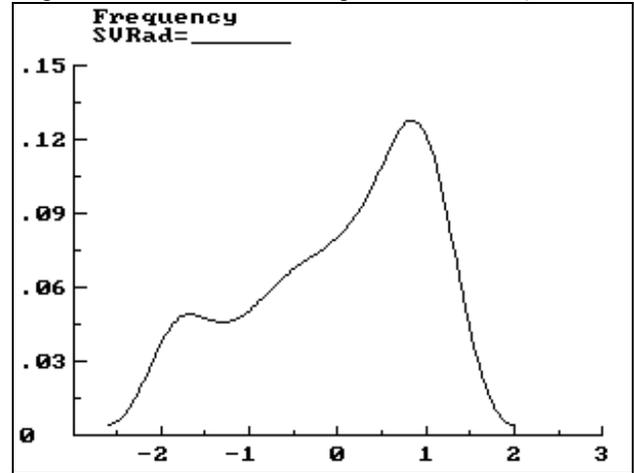


Figura 3: Serie SV de radiación diaria(MJ/m²), (eje x en días)

Figura 4 : densidad SV de radiación diaria,(eje x en MJ/m²)

La figura 5 corresponde al diagrama de dispersión de las variables estudiadas, el cual muestra una clara correlación entre las mismas. Esto dió sentido a la búsqueda de un modelo estructural de correlación entre dichas variables, que permita poder presentar a una de ellas en función de la otra.

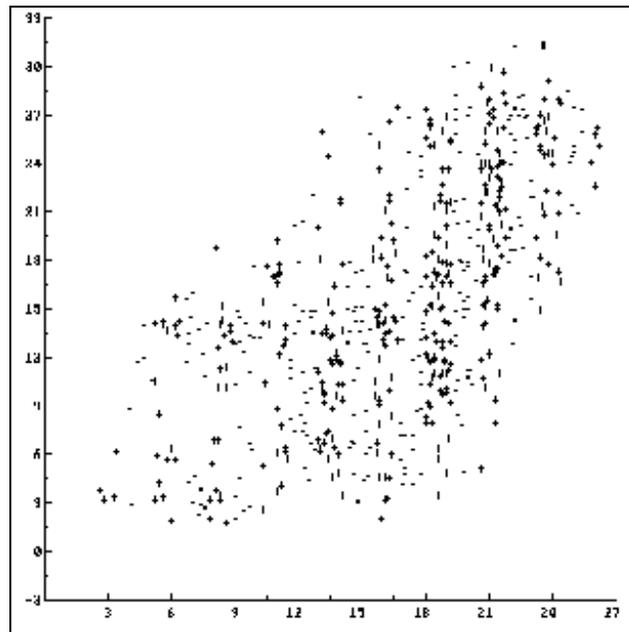


Figura 5 : Diagrama de dispersión entre temperatura y radiación diarias (eje x en °C, eje y en MJ/m²)

ANÁLISIS ESTADÍSTICO DE CORRELACION ENTRE LAS SERIES.

El nexo a través de las diferentes series se realiza mediante las correlaciones de los disturbios que influyen sobre las componentes. En un modelo de factores comunes, algunas, o posiblemente todas, las matrices de covarianza serán de rango reducido. Estos factores comunes son de considerable importancia para el modelado multivariado. Las únicas partes que no aparecen en los modelos univariados son las denominadas matrices de factores y covarianzas, las que son de radical importancia en el contexto de los factores comunes.

En la búsqueda de un modelo de correlación se propusieron diferentes descomposiciones de las variables de estudio, escribiendo a las variables de estudio como la suma de componentes de diversa índole como ser : nivel, tendencia, autorregresión, ciclos e irregular o ruido blanco.

Para poder determinar el modelo óptimo se tuvieron en cuenta las siguientes consideraciones estadísticas teóricas.

En el análisis multivariado, las matrices de covarianzas de los disturbios son calculadas cuando se realiza el proceso de estimación de los hiperparámetros. A continuación se muestra esta matriz, para el análisis bivariado de las series bajo estudio. Notemos que la matriz de covarianza viene dada por la parte triangular superior, correspondiendo los valores de la diagonal principal, a las varianzas de los términos de disturbios en cada serie.

$$\sum_t^* = \begin{pmatrix} 0.063 & -.099 \\ -0.39 & 0.87 \end{pmatrix} \wedge \sum_\eta^* = \begin{pmatrix} 1.464 & 0.936 \\ 0.673 & 0.371 \end{pmatrix}$$

Los elementos indicados en *itálicas* son los correlaciones estimadas. Se observa que la correlación es mucho más grande para los disturbios en los niveles que para la componente irregular, esto significa que tiene sentido establecer la siguiente relación entre las variables bajo estudio, a partir de los respectivos modelos de volatilidad estocástica (SV), ecuaciones (2) y (3)

$$(4) \quad y_{1t}^* = [\mu_{1t} + y_{2t}^*] + \varepsilon_{1t}^* = \mu_{1t} + y_{1t}^*$$

y aplicando la función exponencial miembro a miembro obtenemos la siguiente expresión linealizada :

$$(5) \quad y_{1t} = C \cdot \exp\left\{ \frac{h_{1t}}{2} \right\} \cdot y_{2t} \cdot \varepsilon_{1t}^*$$

donde C es una constante relacionada a la componente de nivel de la primera variable.

A partir de la ecuación (5) se siguieron dos caminos para establecer el modelo de correlación más apropiado :

- 1) sustituyendo y_{2t} por su forma de volatilidad estocástica y aplicando nuevamente logaritmos miembro a miembro, se obtiene la siguiente ecuación :

$$(6) \quad y_{1t}^* = \log y_{1t} = [\log K + (h_{1t} + h_{2t})/2] + \varepsilon_{1t}^* = \mu_{1t} + \varepsilon_{1t}^*$$

en la que K es el producto de las constantes a y c en los modelos (SV), y la expresión entre corchetes representa la componente de nivel. Esta componente queda escrita en función de las caminatas aleatorias h_{1t} y h_{2t} , y en consecuencia el modelo no presenta variables explicativas observadas ; es decir no aparece en forma explícita la variable y_{2t} .

- 2) dejando en forma explícita la variable y_{2t} , y aplicando logaritmos miembro a miembro, se consigue escribir la siguiente ecuación :

$$(7) \quad y_{1t}^* = \log y_{1t} = [\log 3.43 + h_{1t} / 2] + \beta \log y_{2t} + \log \varepsilon_{1t} = \mu_{1t} + \beta y_{2t}^* + \varepsilon_{1t}^*$$

El parámetro β constituye una ponderación sobre la variable independiente y_{2t}^* y cuya estimación se consigue en el proceso de suavizado y filtrado de Kalman. El valor obtenido en este caso es $\beta=0.946$.

Además, si simbolizamos con $(\theta)_E$ al estimador de un parámetro genérico θ , se puede constatar que el estimador $(\log c)_E$ satisface la relación

$$(8) \quad (\log a)_E \cong (\log K)_E - (\log c)_E$$

Los resultados estadísticos complementarios más importantes asociados al modelo óptimo se muestran en la siguiente tabla.

$y = \text{Level} + \text{AR}(1) + 1 \text{ Cycle(s)} + \text{Irregular}$ $y_{_1} : \text{SVRad}$ $y_{_2} : \text{SVTprom}$ Estimation	Resúmen de estadísticos SVRad SVTprom Std.Error 1.730 1.235 Normality 32.00 204.3 H(239) 1.052 1.442
---	--

Modelo con 15 hiperparámetros (0 restricciones). Estimación paramétrica muestral 1. 1 - 720. 1. (T = 720). Núcleo de Log-verosimilitud -1.759383. Convergencia muy fuerte en 96 iteraciones.	<table border="0"> <tr> <td>r(1)</td> <td>0.0051294</td> <td>0.015436</td> </tr> <tr> <td>r(25)</td> <td>0.016204</td> <td>0.034973</td> </tr> <tr> <td>DW</td> <td>1.972</td> <td>1.962</td> </tr> <tr> <td>Q(25,20)</td> <td>13.58</td> <td>30.29</td> </tr> </table>	r(1)	0.0051294	0.015436	r(25)	0.016204	0.034973	DW	1.972	1.962	Q(25,20)	13.58	30.29
r(1)	0.0051294	0.015436											
r(25)	0.016204	0.034973											
DW	1.972	1.962											
Q(25,20)	13.58	30.29											

<p style="text-align: center;">Desviaciones estándar estimadas</p> <table border="0"> <tr> <td>Componente</td> <td>SVRad (razón q)</td> <td>SVTprom (razón q)</td> </tr> <tr> <td>Irr</td> <td>1.3400 (1.0000)</td> <td>0.160066 (1.0000)</td> </tr> <tr> <td>Cy2</td> <td>0.889928 (0.6641)</td> <td>1.0575 (6.6064)</td> </tr> <tr> <td>Ar1</td> <td>0.129650 (0.0968)</td> <td>0.322609 (2.0155)</td> </tr> </table>	Componente	SVRad (razón q)	SVTprom (razón q)	Irr	1.3400 (1.0000)	0.160066 (1.0000)	Cy2	0.889928 (0.6641)	1.0575 (6.6064)	Ar1	0.129650 (0.0968)	0.322609 (2.0155)	<p style="text-align: center;">Parámetros estimados de la componente ciclo</p> <p style="text-align: center;">Matriz de covarianza</p> <table border="0"> <tr> <td>1.202</td> <td>0.40239</td> </tr> <tr> <td>0.57486</td> <td>1.698</td> </tr> </table> <p style="text-align: center;">El coeficiente rho es 0.5841867.</p>	1.202	0.40239	0.57486	1.698				
Componente	SVRad (razón q)	SVTprom (razón q)																			
Irr	1.3400 (1.0000)	0.160066 (1.0000)																			
Cy2	0.889928 (0.6641)	1.0575 (6.6064)																			
Ar1	0.129650 (0.0968)	0.322609 (2.0155)																			
1.202	0.40239																				
0.57486	1.698																				
<p style="text-align: center;">Coeficientes estimados de la componente autorregresiva</p> <p style="text-align: center;">matriz de coeficientes para VAR(1)</p> <table border="0"> <tr> <td>0.98225</td> <td>-0.011537</td> </tr> <tr> <td>-0.0087582</td> <td>0.94630</td> </tr> </table> <p style="text-align: center;">matriz asociada a Ar1</p> <table border="0"> <tr> <td>1.000</td> <td>0.0000</td> </tr> <tr> <td>0.70297</td> <td>1.000</td> </tr> </table>	0.98225	-0.011537	-0.0087582	0.94630	1.000	0.0000	0.70297	1.000	<p style="text-align: center;">Matriz estimada para la covarianza de los disturbios. el triángulo superior derecho muestra correlaciones</p> <p style="text-align: center;">disturbios de Irr</p> <table border="0"> <tr> <td>1.796</td> <td>-0.98410</td> </tr> <tr> <td>-0.21108</td> <td>0.025621</td> </tr> </table> <p style="text-align: center;">disturbios de Cy2</p> <table border="0"> <tr> <td>0.79197</td> <td>0.40239</td> </tr> <tr> <td>0.37867</td> <td>1.118</td> </tr> </table> <p style="text-align: center;">disturbios de Ar1</p> <table border="0"> <tr> <td>0.016809</td> <td>0.28251</td> </tr> <tr> <td>0.011816</td> <td>0.10408</td> </tr> </table>	1.796	-0.98410	-0.21108	0.025621	0.79197	0.40239	0.37867	1.118	0.016809	0.28251	0.011816	0.10408
0.98225	-0.011537																				
-0.0087582	0.94630																				
1.000	0.0000																				
0.70297	1.000																				
1.796	-0.98410																				
-0.21108	0.025621																				
0.79197	0.40239																				
0.37867	1.118																				
0.016809	0.28251																				
0.011816	0.10408																				

Tabla 1 : Resultados estadísticos obtenidos en las etapas de suavizado, filtrado y aplicación del filtro de Kalman

La bondad del ajuste del modelo de correlación puede observarse en las siguientes figuras y tabla :

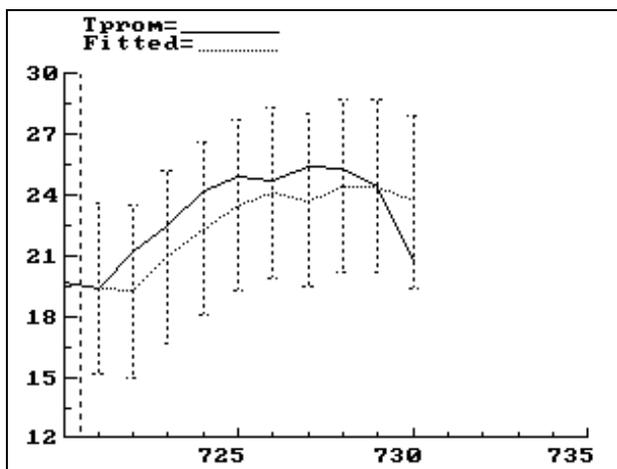


Fig.6 :Predicciones de radiación 10 días hacia adelante (eje y en MJ/m²), (eje x en días)

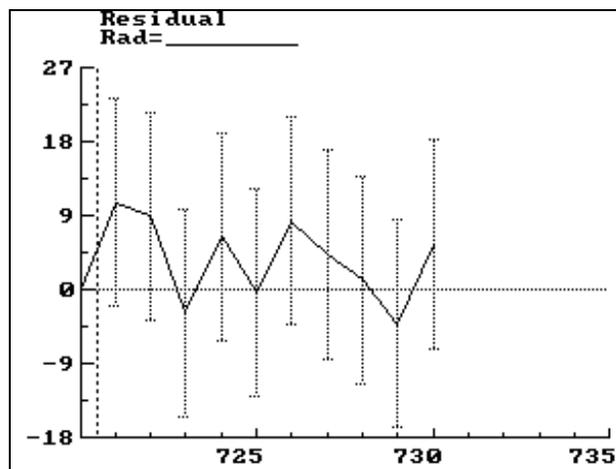
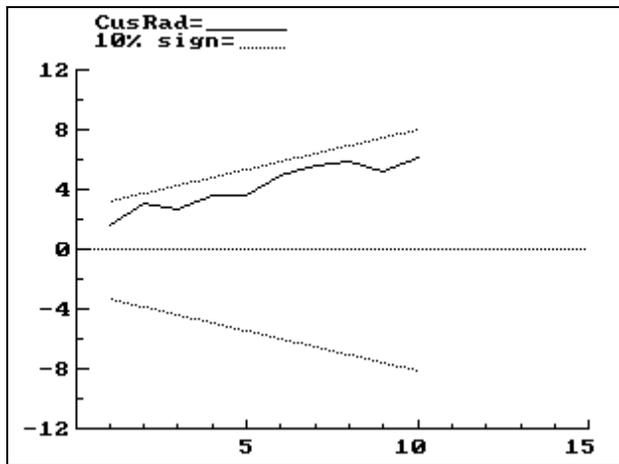


Fig.7 : Comportamiento residual en las predicciones de radiación (eje x en días)



Período	Pronóstico	R.m.s.e.	-Rmse	+Rmse
721. 1	15.953	6.3286	9.6246	22.282
722. 1	16.355	6.9550	9.4001	23.310
723. 1	16.550	7.1094	9.4404	23.659
724. 1	16.665	7.1605	9.5041	23.825
725. 1	16.744	7.1826	9.5617	23.927
726. 1	16.803	7.1951	9.6080	23.998
727. 1	16.845	7.2040	9.6411	24.049
728. 1	16.873	7.2117	9.6610	24.084
729. 1	16.888	7.2191	9.6694	24.108
730. 1	16.895	7.2265	9.6687	24.122
731. 1	16.896	7.2341	9.6618	24.130
732. 1	16.893	7.2420	9.6511	24.135
733. 1	16.889	7.2502	9.6387	24.139
734. 1	16.885	7.2587	9.6260	24.143
735. 1	16.881	7.2672	9.6138	24.148
736. 1	16.878	7.2758	9.6026	24.154
737. 1	16.877	7.2844	9.5924	24.161
738. 1	16.876	7.2930	9.5830	24.169
739. 1	16.876	7.3015	9.5742	24.177
740. 1	16.876	7.3099	9.5658	24.186

Fig. 8 : Comportamiento acumulativo en las predicciones

Tabla 2 : Pronósticos para radiación 10 días hacia adelante

Los estadísticos detallados en la tabla 1, nos permiten avalar los modelos (6) y (7) como válidos, pero preferimos (7) dado que resulta más cómodo su uso desde el punto de vista de las aplicaciones. Este modelo queda expresado en una forma muy similar a las expresiones de correlación del análisis estadístico tradicional. En consecuencia, tomamos como modelo de correlación entre las variables bajo estudio a (trabajando con una confiabilidad del 95%):

$$\log y_{1t} = [\log 3.43 + h_{1t}/2] + 0.946 \log y_{2t} + \log \varepsilon_{1t}, \quad \varepsilon_{1t} \sim \text{IID}(0, 1); \quad h_{1t} = 0.955 h_{1t-1} + \eta_{1t}, \quad \eta_{1t} \sim \text{NID}(0, 0.05) \quad (9)$$

CONCLUSIÓN.

Mediante el empleo del tratamiento estructural de series de tiempo, se consiguió determinar un modelo que permite correlacionar una serie de valores de radiación diaria con otra de valores de temperatura promedio diaria, con una bondad de ajuste muy buena. Esto permite realizar estimaciones de una variable a partir de valores de la otra, a corto plazo y con una confiabilidad muy alta (95%). Un futuro trabajo procurará correlacionar estas variables con los otras tales como por ejemplo, vientos y precipitaciones.

REFERENCIAS.

- Abril J. C. (1999). *Análisis Estadístico de Series de Tiempo Basado en Modelos de Espacio de Estado*. E.U.D.E.B.A. 1999
- Aguilar R. y Collares Pereira M. (1992). Tag : A time dependent, autorregressive, gaussian model for generating synthetic hourly radiation, *Solar Energy*, **49**, 3, 167-174.
- Avila Blas O. J. (1997). Análisis Espectral de Series de Temperatura de Superficie. *Revista FACENA, Univ. Nac. Nordeste*, **13**, 79-99.
- Avila Blas, O., Abril, J.C. y Lesino, G. (1999). *Análisis Estadístico Estructural de Series de Radiación Diaria*, Avances en Energías Renovables y Medio Ambiente, Vol. 3, N° 2, pp. 11.17-11.20.
- Harvey A. C. y Shepard N. (1993). *Estructural Time Series Models*. En *Handbook of Statistics*, Elsevier Science Publishers B.V (Editores), Vol. 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.

ABSTRACT

Following a research line that began in 1998, in this paper a statistical bivariate correlation between diary average radiation and temperature study is done by using the structural analysis methodology, introduced by actual contemporary statisticians. A correlation model between the two variables is presented to generate synthetic diary radiation values from diary temperature values, being able to make forecasting with a high confidential level (95%).

