ANALISIS ESTADISTICO ESTRUCTURAL DE SERIES DE RADIACION DIARIA

O. J. Avila Blas[1], J. C. Abril[2], G. Lesino[3]
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

En el presente trabajo se realiza un estudio estadístico de una serie de radiación promedio diaria, empleando la metodología del análisis estructural, el que ha cobrado desde hace poco tiempo, gran relevancia desde el punto de vista de poseer muy buenas características y poder ser aplicado a situaciones antes no siempre solucionables con los modelos ARIMA de Box y Jenkins. Se presenta un modelo para poder generar sintéticamente valores de radiación diarios, que permite realizar pronósticos con una alta confiabilidad.

.

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%. En muchos casos esto es suficiente ya que en la situación analizada (por ejemplo, la electrificación de una escuela aislada mediante un pequeño sistema fotovoltaico) son grandes las diferencias de costos con otras alternativas o bien, existe un monto fijo de dinero para realizar el proyecto. 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. Actualmente se trabaja con el “Año Meteorológico Típico”, usado por los solaristas, pero éste adolece de algunas fallas tales como la arbitrariedad en la asignación de factores de peso a los valores medios y la pérdida de algunas correlaciones. Este problema se agudiza cuando se trata de valores horarios (calculados a partir de datos diarios) y para series de días, como en el caso de sistemas con inercia como las viviendas solares y en general los sistemas de acumulación para varios días. Ultimamente se han propuesto métodos estadísticos de generación de series de algunas variables de interés : temperatura de superficie (Avila Blas, 1997), radiación horaria (Aguiar, 1992), pero siempre utilizando modelos basados en el tratamiento de Box y Jenkins, es decir, modelos ARIMA(p,d,q). No se conocen hasta el momento trabajos de modelado de series de tiempo con la metodología del enfoque estructural introducido por estadísticos contemperáneos (Harvey y Shepard, 1993 ; Abril, 1997). La idea central de este trabajo en iniciar este tipo de estudio muy nuevo, al caso de una serie de valores de radiación promedio diaria, que corresponde a mediciones realizadas entre los años 1981 y 1996 en la localidad de Cerrillos (Provincia de Salta).

ANALISIS ESTADISTICO ESTRUCTURAL.

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.

La metodología de trabajo en la selección para los modelos denominados estructurales, es diferente en muchos aspectos al tratamiento clásico de Box y Jenkins de los modelos **ARIMA(p,d,q)**. Se pone menos énfasis en la observación del correlograma de diversas transformaciones de la serie con el objeto de obtener una especificación inicial. En lugar de ello, el énfasis está puesto en la formulación del modelo en términos de componentes cuya presencia está sugerida por el conocimiento del fenómeno bajo estudio, de sus aplicaciones o por una inspección del gráfico de la serie original. Una vez que el modelo ha sido estimado, el mismo tipo de tests de diagnóstico para los modelos ARIMA puede ser aplicado. Además el estudio se completa con tests de falta de normalidad y heterocedasticidad (varianza variable en el tiempo), tests para la calidad predictiva en períodos posteriores a la

[1] Probabilidades y Estadística, Depto. de Matemática, Fac. Cs. Exactas - UNSa
[2] INIE, Instituto UNT - CONICET
[3] INENCO, Instituto UNSa - CONICET

muestra y gráficos de los componentes suavizados. Además no es necesario que la serie sea estacionaria, es decir, con media y varianza constante en el tiempo, lo que sucede muy a menudo en variables de clima.

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 suaviazador 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. Otra importancia de este enfoque es que los modelos estructurales pueden ser escritos, mediante transformaciones adecuadas, como modelos ARIMA, ésta es la denominada forma reducida. La representación matemática de un modelo de espacio de estado relaciona al vector de disturbios $\{\ _t\}$ con el vector de observaciones $\{y_t\}$ a través de un proceso de Markov $\{\ _t\}$ y es :

$$y_t = Z_t \ \ _t + G_t \ \ _t , \qquad _t \sim N(0, H_t)$$
$$_t = T_t \ \ _{t-1} + H_t \eta_t , \quad \eta_t \sim N(0, Q_t) ; \quad _0 \sim N(\boldsymbol{a_0}, \boldsymbol{P_0}) , \quad t=1,\dots,T$$

donde $_t$ es el vector de estado de orden mx1, $_t$ es un vector de disturbios de orden kx1 y las matrices del sistema Z_t, T_t, G_t, y H_t tienen dimensiones Nxm, mxm, Nxk y mxk 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-Ljong) 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.

El filtro de Kalman tiene como objetivo actualizar nuestro conocimiento del sistema cada vez que una nueva observación y_t es obtenida. Cuando el modelo ha sido puesto en forma de espacio de estado, se pueden aplicar una gran cantidad de algoritmos, siendo el punto central el mencionado filtro. Este filtro es un procedimiento recursivo que permite computar el estimador óptimo del vecto de estado en el momento t, basado en la información disponible hasta el tiempo t, la que incluye a y_t. Cuando el supuesto de normalidad se deja de lado, no hay ninguna garantía de que el filtro produzca la media condicional del vector de estado. De todos modos es todavía un estimador óptimo en el sentido de que minimiza el error medio cuadrático medio dentro de la clase de todos los estimadores lineales. El otro paso importante en este procedimiento es el denominado "suavizado", el que es una recursión hacia atrás en la que se emplean las inversas de las matrices relacionadas al proceso de filtrado. Las magnitudes suavizadas juegan el importante rol de pivotes en la construcción de tests de diagnóstico para observaciones atípicas y cambios estructurales. 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.

La serie original tratada aquí consta de T=5842 valores diarios. Pero, dada la ventaja del tratamiento estructural, podemos trabajar con una serie sustancialemente más pequeña, por ejemplo tomando el último año, esto es 1996 ; en particular los 350 primeros valores, a fin de poder compararlos con los datos predichos por la serie modelada. La representación gráfica de esta serie junto con el correspondiente correlograma se presentan en las figuras 1 y 2 :

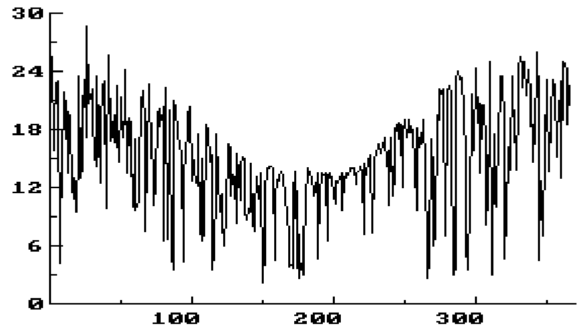

Figura 1 : Serie de radiación diaria (MJ/m^2), (eje *x* en días)

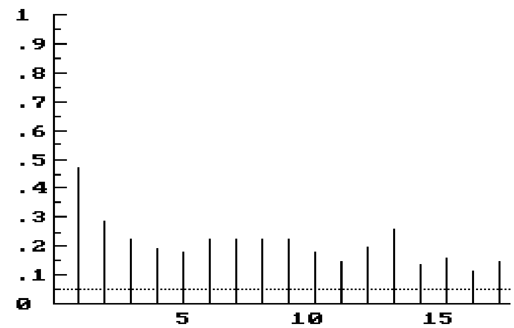

Figura 2 : correlograma de la serie de radiación

ANALISIS ESTADISTICO DE LA SERIE.

Antes de llegar al modelo final, se han intentado diferentes modelos como ser : tomar la variable original y considerarla como una combinación lineal de componentes de nivel e irregular estocásticas ; ó bien, realizarle una diferenciación de orden 1 y proponer la misma descomposición anterior, y otros , pero en todos ellos el modelado no resultó con las propiedades deseadas, dado que los valores de los estadísticos asociados a la bondad de ajuste : BS, Q y DW no fueron los adecuados. La transformación final consiste en proponer un **modelo de varianza o volatilidad estocástica** (SV), el que puede formularse como sigue :

$$y_t = \sigma_t\, \varepsilon_t = \sigma\, \varepsilon_t \exp(h_t/2)\,, \quad {}_t \sim \mathrm{IID}(0, 1), \quad t=1,\ldots,T$$
$$h_t = \phi\, h_{t-1} + \eta_t \;, \quad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2), \qquad |\phi| \quad 1$$

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.

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 :

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

donde s_y^2 es la varianza de los datos. Para poder realizar el modelado, se le aplica dicha transformación a la serie original ; la serie resultante (y_t*) y su correspondiente correlograma se presentan en las figuras 3 y 4 :

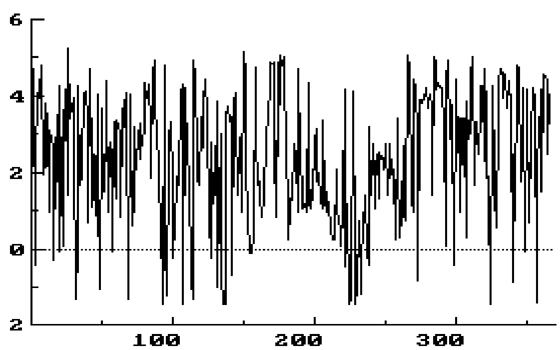

Figura 3 : serie y_t* (log MJ/m^2), (eje *x* en días)

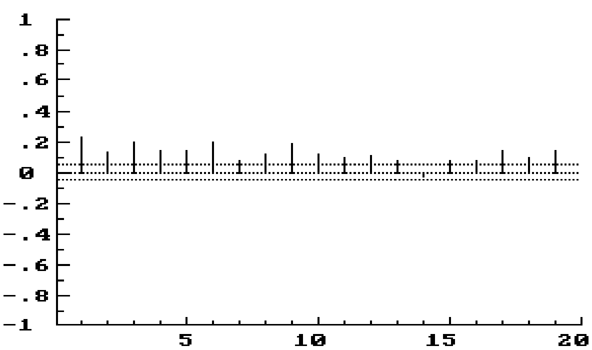

Figura 4 : correlograma de la serie y_t*

y luego se la ajusta a un modelo con componentes : nivel fijo + AR(1) + irregular. Algunos de los resultados computacionales usando el soft STAMP 5.0 (Koopman *et al*., 1995) se muestran a continuación :

Resumen de estadísticos para la variable y_t*

Error est.	1.638
Normalidad	15.06
H(121)	0.88120
r(1)	0.053829
r(18)	0.0051983
DW	1.891
Q(18,16)	19.62

Ec. 2 : Deviaciones estándar estimadas de los disturbios.

Componente	y_t* (razón q)
Irr	1.5456 (1.0000)
Ar1	0.233212 (0.1509)

Ec. 1 : Coeficiente autorregresivo estimado

El coeficiente rho del AR(1) es ϕ=0.9595499.

Notemos que el estadístico de normalidad no es muy bueno, lo que es inevitable dado que el modelo transformado no es Gaussiano ; por ello realizamos la estimación suavizada del proceso de volatilidad, h_t que permite obtener una bondad de ajuste muy buena.

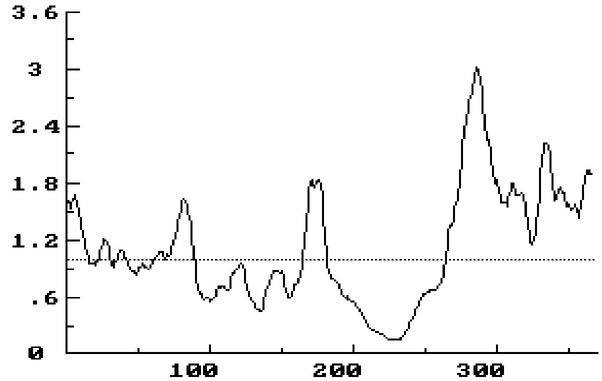

Figura 6 : estimación suavizada del componente h_t , (eje *x* en días)

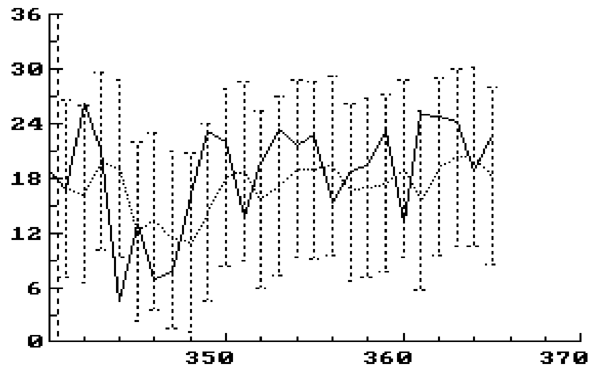

Figura 7 : Bondad de ajuste de la serie y_t* vs. la ajustada

Para estimar el parámetro σ^2 , se forma la serie de observaciones corregidas por heterocedasticidad y a ella se le computan los estadísticos descriptivos. Finalmente, el modelo que permite ajustar nuestros datos se puede escribir como :

$$y_t = \sigma_t \varepsilon_t = 3.43\, \varepsilon_t \exp(h_t/2)\ , \quad {}_t \sim IID(0, 1), \quad t=1,\ldots,365\ ; \quad \text{donde} \quad h_t = 0.955\, h_{t-1} + \eta_t\ , \quad \eta_t \sim NID(0\ , 0.05)$$

A continuación se muestra la predicción correspondientes a 15 días hacia adelante, empleando la serie modelada

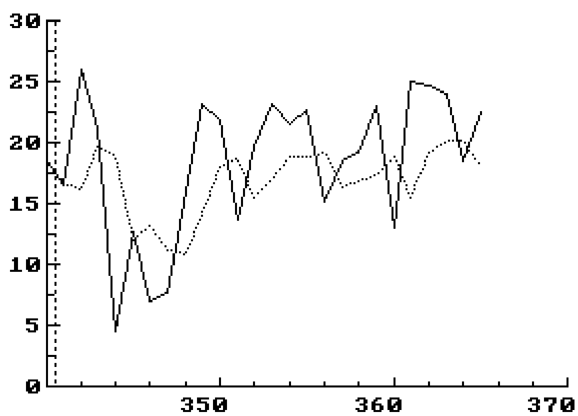

Figura 8: comparación entre los 15 últimos datos originales y los predichos por la serie ajustada (línea débil) (MJ/m^2) , (eje *x* en días)

Ec. 1 : Pronósticos para Anti-log (y_t*)

Período	Pronóst.	R.m.s.e.	- Rmse	+ Rmse
351. 1	16.899	91.170	-69.271	113.07
352. 1	17.367	90.039	-68.671	111.41
353. 1	14.869	88.923	-68.054	109.79
354. 1	16.401	87.827	-67.426	108.23
355. 1	16.963	86.755	-66.792	106.72
356. 1	18.051	85.708	-66.157	105.26
357. 1	18.164	84.688	-65.524	103.85
358. 1	16.799	83.696	-64.897	102.50
359. 1	17.456	82.734	-64.277	101.19
360. 1	15.133	81.801	-63.668	99.934
361. 1	16.828	80.898	-63.070	98.726
362. 1	17.540	80.026	-62.486	97.566
363. 1	18.068	79.183	-61.915	96.451
364. 1	17.011	78.370	-61.358	95.381
365. 1	16.769	77.586	-60.817	94.354

CONCLUSIÓN.

Mediante el empleo del tratamiento estructural de series de tiempo, se consiguió modelar una serie de valores de radiación diaria, con una bondad de ajuste muy buena. Esto permite realizar estimaciones a corto plazo con una confiabilidad alta. Un futuro trabajo procurará correlacionar estos modelos con los correspondientes a otras variables como ser temperatura, heliofanía y vientos.

REFERENCIAS.

Abril J. C. (1999). *Análisis Estadístico de Series de Tiempo Basado en Modelos de Espacio de Estado*. E.U.D.E.B.A. En prensa.

Aguiar 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.

Harvey A. C. y Shepard N. (1993). Estructural Time Series Models. En *Handbook of Statistics*, Elsevier Sciencie 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.