

MODELADO ESTRUCTURAL DE SERIES DE VELOCIDAD DE VIENTO EN LA CIUDAD DE SALTA: CONTROL ESTADÍSTICO DE UN MODELO PREVIO

Orlando J. Avila Blas, Judith Franco, Nahuel Salvo

Facultad de Cs. Exactas, Universidad Nacional de Salta, Avenida Bolivia 5150, 4400 Salta, Argentina
Tel.: 54-387-4255385 – Fax: 54-387-4255449

E-mail: oblas@unsa.edu.ar, francoj@unsa.edu.ar, nahuel@unsa.edu.ar

RESUMEN – En este trabajo se realiza el control estadístico (testeo) de un modelado estadístico estructural previamente diseñado para una serie temporal de valores promedio diarios de la variable velocidad de viento. Esto se convalida, introduciendo un algoritmo relacionado a un modelo de regresión dinámica de segundo orden, enmarcado dentro de la distribución de Poisson. Este nuevo modelo permite realizar el análisis estadístico de series de tiempo no gaussianas, lo que le otorga flexibilidad al mismo realizando pronósticos con una confiabilidad muy alta (95%).

Palabras claves : velocidad de vientos, energía eólica, estadística, modelado estructural, pronósticos.

INTRODUCCION

Para el diseño de sistemas eólicos, es necesario conocer las velocidades y direcciones de vientos, con una cierta confiabilidad. Si bien existen mapas del recurso eólico, éstos están confeccionados a una escala muy grande y no tienen mucha precisión. A partir de los datos medidos en distintas estaciones meteorológicas se puede realizar un tratamiento estadístico que involucre el diseño de modelos para generar sintéticamente valores promedios y estudiar variables correlacionadas entre sí y en el tiempo, en particular empleando la metodología del análisis estructural de series de tiempo.

En este sentido se pueden tomar como referencia, trabajos en los que se presentan métodos estadísticos de generación sintética de series de otras variables de interés en problemas físicos : temperatura de superficie (Avila Blas, 1997), radiación horaria (Aguiar, 1992), utilizando modelos basados en el tratamiento de Box y Jenkins, es decir, modelos ARIMA(p,d,q). En particular, los siguientes trabajos realizan el modelado de series de tiempo con la metodología del enfoque estructural, introducido por estadísticos contemporáneos (Abril, 1999) y se refieren a variables tales como radiación promedio diaria, temperatura de superficie promedio diaria y correlaciones entre ellas (O. Avila Blas *et al.*, 1999; O. Avila Blas *et al.*, 2000). También es importante nombrar un trabajo referido al tratamiento estadístico de velocidad de vientos en la ciudad de Salta (O. Avila Blas *et al.*, 2002), que se tomará como base para el presente artículo.

En todas las referencias mencionadas se logró "levantar" la condición de normalidad en la distribución de las variables de estudio, lo que le brindó una mayor robustez estadística a las variables, al mismo tiempo que una flexibilidad en su tratamiento, fundamentalmente desde el punto de vista de efectuar pronósticos con una alta confiabilidad (entre 95% y 99%). Este resultado es muy importante ya que en la mayoría de las situaciones físicas de estudio actual, la distribución empírica de las variables bajo estudio, no son normales o gaussianas, y realizar este tipo de supuestos conduce en general a conclusiones poco confiables o poco prácticas.

METODOLOGIA ESTADISTICA EMPLEADA

El núcleo de la teoría de los modelos estructurales de series de tiempo, es que éstos pueden ser expresados como modelos de regresión en donde las variables explicativas son funciones del tiempo, con coeficientes que pueden cambiar a través del tiempo, lo que le otorga al modelo la propiedad de ser dinámico. La estimación actual de los coeficientes, se logra poniendo al modelo en forma de espacio de estado y aplicándole luego el denominado Filtro y Suavizador de Kalman (SFK). Se emplean algoritmos específicos para hacer estimaciones, predicciones y para los suavizados. Esto último significa computar el 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, entre los que figuran: varianzas asociadas a ruidos blancos de la serie, y los coeficientes del modelo de regresión dinámica. Estos pueden ser estimados por el método de máxima verosimilitud, construyendo la función específica a optimizar, que en la gran mayoría de los casos se consigue a partir exclusivamente de los datos que conforman la serie bajo estudio (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 (uno de los objetivos principales de este artículo). 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, estacionalidad, ciclos, y para el cálculo de estadísticos que son empleados para detectar observaciones atípicas (“outliers”), cambios estructurales y tratamiento de datos con distribuciones empíricas multimodales.

El modelo matemático general 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 es :

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

donde α_t es el denominado “vector de estado”, ε_t es el “vector de disturbios o ruido blanco” y las matrices Z_t , T_t , G_t y H_t tienen la propiedad que sus elementos son funciones del tiempo. Los disturbios están mutuamente no correlacionados con medio cero y varianza H_t . 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.

Para el caso que tratamos en este artículo, no podemos suponer normalidad en los disturbios, por lo que es necesario conseguir una transformación matemática apropiada, a fin de encontrar un modelo similar al de la ecuación (1), en el sentido que los disturbios transformados tengan una distribución aproximadamente gaussiana. Solamente así se puede emplear la metodología estructural a fin de poder obtener una modelización confiable y robusta.

ANÁLISIS ESTADÍSTICO DE LA SERIE

La base de datos analizada está formada por valores promedios diarios de velocidad de vientos en la ciudad de Salta (medidos en *Km/h*) y abarca el período comprendido desde 1982 hasta 1990. En un trabajo previo, se realizó el modelado estadístico estructural tomando una serie más corta (O. Avila Blas *et al*, 2002), la de valores promedios diarios para los años 1982 y 1983, y se generalizó, empleando la metodología estadística estructural, para la serie correspondiente al período 1982-1988. Para realizar el control o convalidación de este modelo, se procedió a anexar a la misma, los valores promedios registrados para los años 1989 y 1990. Cuando se realiza el correspondiente estudio estadístico, se descubre la existencia de un período de 2 años, lo que nos habilita, a los fines de la validación, analizar solamente la serie para los años 1989-1990 ($T=719$ datos), pero manteniendo presente la estructura estadística diseñada para la serie de 1982-1988.

El período 1989-1990 tiene un comportamiento muy particular, como puede verse en la Figura 1, por ejemplo, se observan datos atípicos o "outliers", los que posiblemente corresponda a situaciones físicas reales observadas (como por ejemplo, una ráfaga) o bien, datos mal medidos.

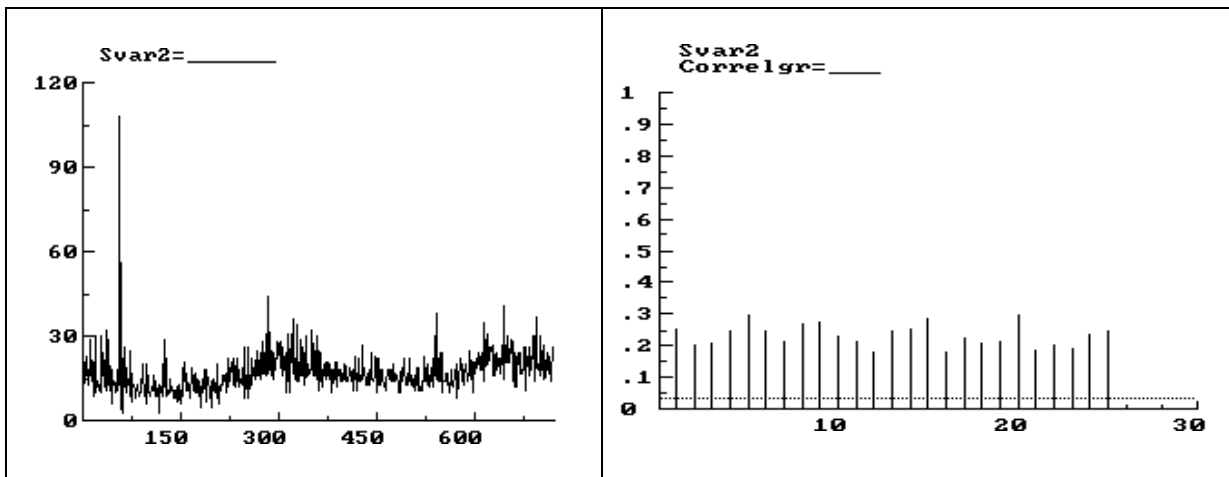


Fig. 1: Serie de vientos y_t (Svar2) (en *Km/h*) (eje x: número de día)

Fig.2: Correlograma de la serie bajo estudio

Para el tratamiento estadístico de la variable en cuestión (1989-1990) se denota y_t ="Svar2". Nota: Cuando en una gráfica no se indiquen unidades es porque se trata de un índice o estadístico adimensional.

Es importante destacar que la serie bajo estudio tiene asociado un correlograma (Figura 2) que muestra una convergencia muy lenta, lo cual también se puede atribuir a un ciclo subyacente no detectable con la observación de la figura 1 ó bien a la presencia de datos atípicos.

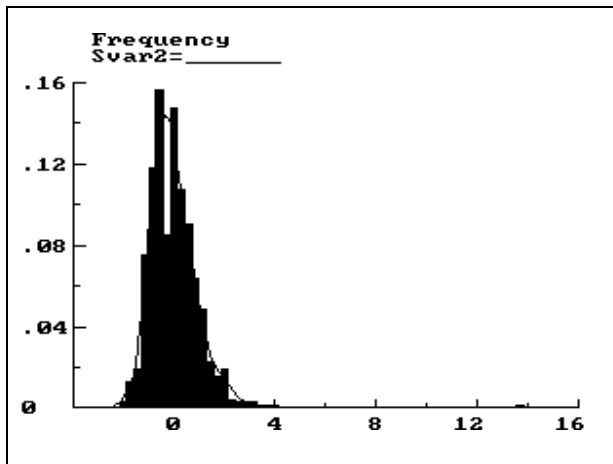


Fig. 3: Distribución empírica de la serie bajo estudio

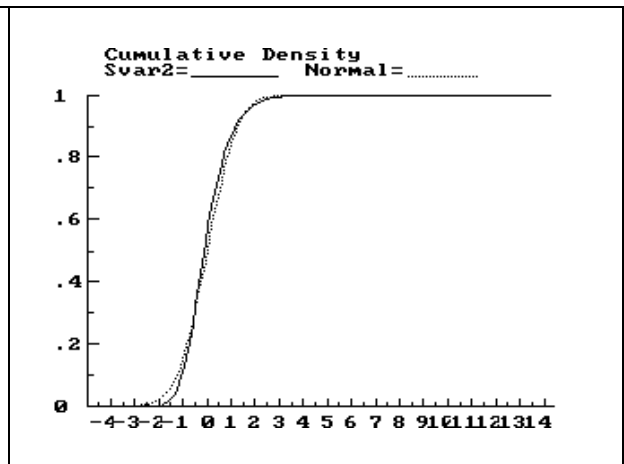


Fig. 4: Distribución acumulativa de la serie bajo estudio

Por otra parte, la distribución empírica de la serie observada resulta claramente no gaussiana, más bien se observa una clara asimetría por la derecha y es de tipo bimodal, tal como puede apreciarse en la Figura 3. Este hecho también se manifiesta en el contraste de la distribución acumulativa empírica con la distribución normal o gaussiana, de la Figura 4, la cual refleja un no ajuste significativo. El hecho de observar una pesada cola a derecha de la distribución (Figura 3), llevó a pensar que la resolución del problema está enmarcado dentro de la familia de distribuciones exponenciales, en particular la distribución de Poisson. Observaciones similares pueden hacerse luego del proceso de análisis residual, lo que se puede mostrar en las Figuras 5, 6 y 7, correspondientes a la serie asociada residual, es decir, la serie de los $\{\varepsilon_t\}$ presentados en la ecuación (1).

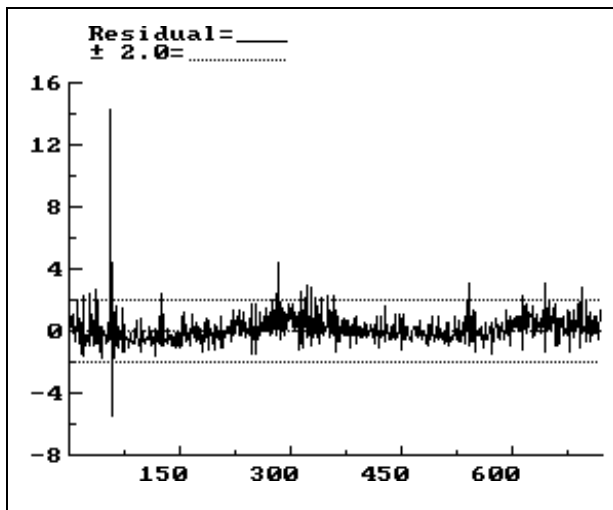


Fig. 5: Serie residual asociada a la serie bajo estudio

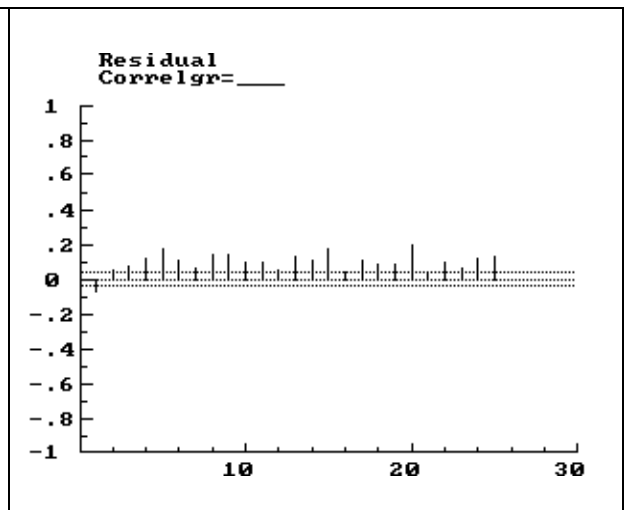


Fig. 6: Correlograma residual asociado a la serie de estudio

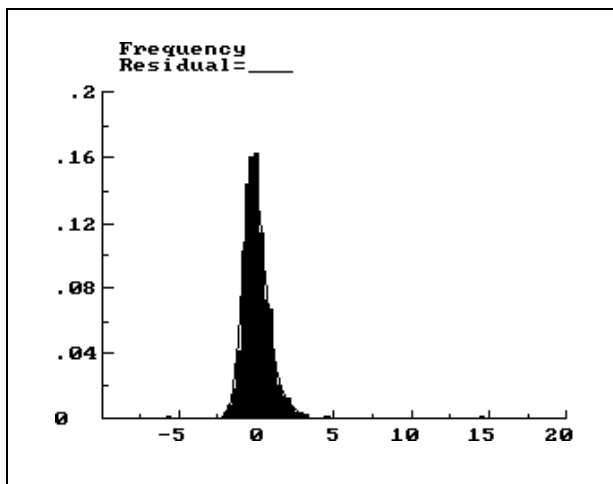


Fig.7: Distribución empírica de la serie residual

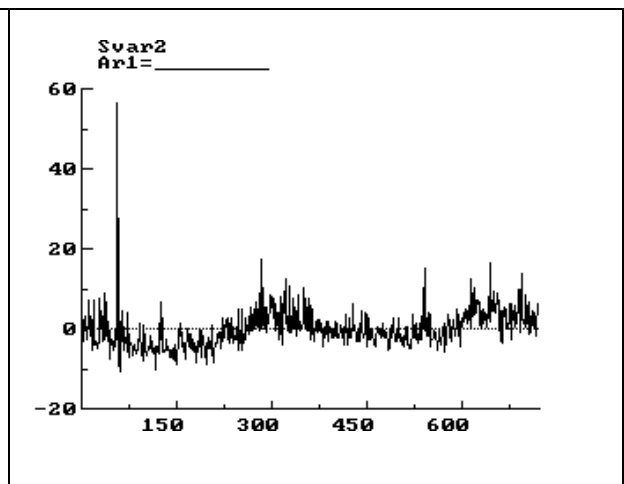


Fig.8: Componente autorregresiva AR(1) para la serie y_t

En particular, en la Figura 8, se presenta el comportamiento de la serie definida por la caminata aleatoria de la segunda igualdad de la ecuación (1), la que presenta un aspecto muy similar a la serie bajo estudio, indicando que la componente autorregresiva tiene peso considerable en la definición del modelo final que se pretende obtener y convalidar. El modelo estadístico estructural desarrollado en el trabajo “Modelo estadístico estructural de series de velocidad de vientos en la ciudad de Salta”, publicado en el Vol. 10 de la ERMA, quedó enmarcado dentro de una

Regresión Dinámica con observaciones de tipo Poisson

La teoría general para el tratamiento de observaciones de la familia exponencial, puede verse con más detalle en (J. C. Abril, 2000), y está basado en técnicas iterativas para computar los estimadores de los parámetros asociados al modelo, desarrolladas por (Durbin y Koopman, 1997: 2000: 2001) quienes emplearon algoritmos rápidos de suavizado y filtrado de Kalman, y también métodos para encontrar estimadores aproximados por máxima verosimilitud (EM).

Para la serie de vientos estudiada previamente, se tomaron aspectos de esta teoría y se diseñó un modelo de regresión que tiene un solo regresor z_t con coeficiente α_t que varía en el tiempo de acuerdo a un camino aleatorio de la forma

$$\alpha_t = \alpha_{t-1} + \eta_t \text{ con } \eta_t \sim N(0, \sigma^2) ; t=1,2,\dots,T \quad (2)$$

Las observaciones y_t son variables con distribución de Poisson con medias $Exp(z_t \alpha_t)$. Esta es una forma simple de modelo loglineal y es un caso particular de un modelo más general, con la diferencia de que aquí el coeficiente de regresión varía en el tiempo. Estas consideraciones permitieron obtener un modelo estadístico estructural con muy buenas propiedades, y con una alta confiabilidad para poder realizar pronósticos (95%).

Ahora, al haber incorporado nuevos datos a la serie original, y de la observación del comportamiento de la serie resultante, en particular, la información que brinda el correlograma (Figura 2), la función de distribución empírica (Figura 3), y los coeficientes característicos descriptivos: índice de normalidad, variabilidad, asimetría y kurtosis, se hace necesario proceder a testear el modelo (2) usando la nueva serie. Este tipo de estudio se realiza planteando a partir de la función de densidad conjunta del vector de estado y los nuevos datos, algún tipo de proceso iterativo (en general similar al involucrado en el tratamiento de la serie previa), y buscando una modelización final que siga perteneciendo a la familia de la serie original (en nuestro caso, la familia Poisson). Para ello, y luego de intentar diferentes modelos iterativos, se logra el óptimo que se describe a continuación.

Modelo de Regresión Dinámica de Orden 2, con observaciones de tipo Poisson generalizada

Sin pérdida de generalidad, y a fin de poder hallar el modelo cuasi-normal análogo al definido por (1), podemos tomar para compensar el proceso de inicialización, al vector de estado inicial α_0 como una constante conocida, y en consecuencia, el logaritmo de la función de densidad conjunta de α e Y_n , tiene la forma

$$\log p(\alpha, Y_n) = \frac{-1}{2\sigma^2} \sum_{t=1}^T (\alpha_t - \alpha_{t-1})^2 + \sum_{t=1}^T \left[\left(\frac{\pi_t}{1-\pi_t} \right)^2 \cdot z_t \alpha_t y_t - \pi_t \exp(z_t \alpha_{t-1}) - \log(y_t!) \right] \quad (3)$$

En la que el factor de peso $\left[\left(\frac{\pi_t}{1-\pi_t} \right)^2 \right]$ es el que caracteriza el orden de la regresión dinámica, y $\pi_t = \frac{z_t \alpha_t}{1+z_t \alpha_t}$ es una transformación de tipo logística.

Diferenciando con respecto a α e igualando a cero, obtenemos los estimadores EM para $\alpha_1, \dots, \alpha_T$ como solución de las ecuaciones no lineales

$$\frac{1}{\sigma^2} (\alpha_{t-1} - 2\alpha_t + \alpha_{t+1}) + \pi_t z_t [y_t - \exp(z_t \alpha_t)] = 0 \quad (4)$$

$$t=1,2,\dots,T-1$$

con el primer término reemplazado por

$$\sigma^{-2} (\alpha_{n-1} - \alpha_n) , \text{ para } t=T \quad (5)$$

Para poder resolver (4) junto con (5), se procede a considerar el modelo Gaussiano análogo, el que se obtiene como en la mayoría de los casos no normales tratados en (Abril, 2000). Este tiene el mismo camino aleatorio que el dado por la ecuación (2), pero en lugar del modelo loglineal, tenemos el modelo de regresión lineal

$$y_t = \pi_t z_t \alpha_t + \varepsilon_t , \text{ con } \varepsilon_t \sim N(0, \sigma_t^2) ; t=1,2,\dots,T \quad (6)$$

el cual es un modelo gaussiano de regresión por el origen, con un coeficiente que varía en el tiempo, y ponderado por la

fracción logística π_t y con disturbios heterocedásticos (varianza no constante en el tiempo) y cuyo logaritmo de la función de densidad conjunta viene dado por

$$\log p(\alpha, Y_n) = -\frac{1}{2\sigma^2} \sum_{t=1}^T (\alpha_t - \alpha_{t-1})^2 + \sum_{t=1}^T (y_t - z_t \alpha_t)^2 \quad (7)$$

el cual, diferenciando e igualando a cero, nos da para el EM al resolver las ecuaciones

$$\frac{1}{\sigma^2} (\alpha_{t-1} - 2\alpha_t + \alpha_{t+1}) + \frac{z_t}{\sigma_t^2} [y_t - z_t \alpha_t] = 0, \text{ con } \sigma^{-2} (\alpha_{n-1} - \alpha_n) \text{ para } t=T$$

$$t = 1, 2, \dots, T-1 \quad (8)$$

A fin de poder resolver las ecuaciones dadas por (4) y (5), las linealizamos poniéndolas de la misma forma que en (8), para luego resolver las ecuaciones resultantes mediante el SFK. Se procedemos ahora de manera similar que en (O. Avila Blas *et al*, 2002) tomando $\hat{\alpha}_t$ como un valor experimental de α_t , y expandimos alrededor del valor dado, obteniendo

$$y_t > = y_t - \exp(z_t \hat{\alpha}_t) (1 - z_t \hat{\alpha}_t)$$

$$z_t > = \exp(z_t \hat{\alpha}_t) z_t \quad (10)$$

$$\sigma_t^2 > = \exp(z_t \hat{\alpha}_t)$$

De este modo el segundo sumando en la derecha de (4) se convierte en

$$\frac{z_t >}{\sigma_t^2 >} (y_t > - z_t > \hat{\alpha}_t) \quad (11)$$

el que tiene la misma forma que el segundo término de (8), y en consecuencia podemos resolver la ecuación linealizada, y a partir de ésta obtener una mejor aproximación a los estimadores $\hat{\alpha}_t$ para el modelo Poisson, aplicando exactamente el mismo algoritmo del SFK que se usa para la estimación de los correspondientes al modelo Gaussiano (2) y (6).

Los valores resultantes son sustituidos en (8) para obtener nuevos valores de los $y_t >$, $z_t >$ y $\sigma_t^2 >$, los que a su vez son tratados con el SFK y así sucesivamente hasta lograr la convergencia deseada. Cualquier método de inicialización considerado en los modelos gaussianos pueden ser empleados a ahora en adelante. Esta metodología se aplicó a nuestra serie de vientos, tomando como criterio de convergencia para el vector de estado, parar cuando el cambio relativo promedio en el EMP para todos los elementos del estado para $t=1,2,\dots,T$ era menor que 10^{-7} .

ANÁLISIS DE LOS RESULTADO

El criterio para realizar la estimación de los hiperparámetros fue el de parar el proceso iterativo cuando el cambio relativo promedio de los hiperparámetros fuese menor que 10^{-10} , habiendo iniciado el proceso de estimación de los hiperparámetros por el método del filtro de Kalman extendido. El proceso recursivo conduce a la definición de una variable transformada final, denominada "Svar22". Se estudia el comportamiento estadístico de esta variable y luego se le aplican a la misma, el conjunto de transformaciones inversas asociadas a fin de poder encontrar la forma sugerida por la ecuación (6). Con ello, se halla el modelo final (óptimo) para la variable y_t (el soft la renombra como "Svar1") que nos permite realizar la convalidación del modelo en cuestión, y así poder generar sintéticamente valores promedios, los que son pronósticos con alta confiabilidad para la variable bajo estudio. Los resultados computacionales más importantes obtenidos se muestran en la siguiente Tabla 1.

y _t =Svar1= Nivel+ AR(1) + 1 ciclo + Irregular (modelo óptimo)				Resúmen de estadísticos para Svar1	
Coeficientes estimados del vector de estado final:				Error estándar	5.918
Variable	Coefficient	R.m.s.e.	t-value	Normalidad	2.167e+005
Lvl	16.954	2.5659	6.6077 [0.0000] **	H(239)	0.39937
Cy1_1	0.00000	0.00000		r(1)	0.0012299
Cy1_2	0.00000	0.00000		r(25)	0.045725
Ar1	3.0582	2.7943		DW	1.994
Convergencia fuerte en 43 iteraciones. (likelihood cvg 2.497375e-016)				Q(25,20)	36.52
				R ²	0.95424
				El coeficiente rho de AR(1) es 0.9954262	

Tabla 1: resultados descriptivos más importantes para la variable transformada Svar1

En esta tabla podemos ver que la variable y_t se puede expresar como la suma de 4 componentes: un nivel fijo (media fija), una caminata aleatoria de orden 1, 1 componente cíclica y el ruido blanco. En particular, la prueba t para el nivel, que es muy significativa, ajusta apropiadamente al hecho que el modelo final debe cumplir con la propiedad de estacionariedad. Los estadísticos de prueba son muy significativos, observar en particular el valor del coeficiente de determinación R^2 , y el coeficiente ρ de la componente autorregresiva, que es muy cercano a 1 (condición que indica que los pronósticos a efectuar con el modelo son robustos). El proceso de maximización de la verosimilitud es fuertemente convergente en 43 iteraciones. Gráficamente, se puede observar en la Figura 9 el modelo óptimo ajusta muy apropiadamente, al contrastar los 44 últimos valores medidos (Svar1, en línea continua), con los generados sintéticamente (Fitted, en línea débil), ya que éstos caen dentro de los respectivos intervalos de confianza.

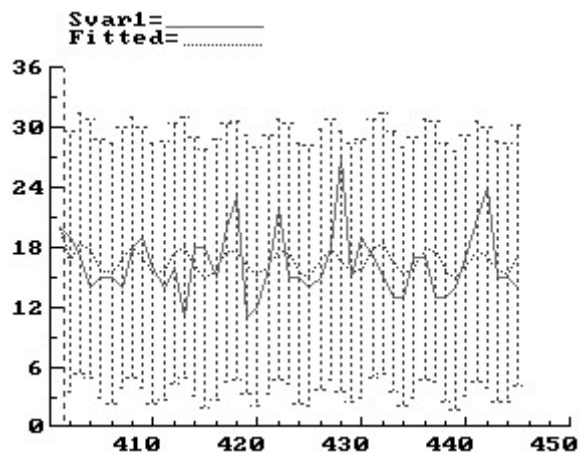


Figura 9: comparación entre los 44 últimos valores observados y los pronosticados por el modelo óptimo

CONCLUSIÓN

Mediante el empleo del tratamiento estructural de series de tiempo y empleando una estructura matemática basada en la distribución de Poisson, en particular un modelo dinámico de segundo orden, se consiguió convalidar (testear) un modelo previo, a partir de una serie de valores promedios diarios de velocidad de viento, con una bondad de ajuste muy buena. Esto permite generar sintéticamente valores de la variable en cuestión y realizar pronósticos hasta 15 días hacia delante, con una confiabilidad muy alta (95%). Este artículo introduce una nueva perspectiva en cuanto a las metodologías que se pudieron estar utilizando hasta el momento, con el fin de optimizar el uso de series temporales de valores generadas en forma sintética.

REFERENCIAS

- Abril J. C. (2000). *Análisis Estadístico de Series de Tiempo Basado en Modelos de Espacio de Estado*. 1º Edición. E.U.D.E.B.A.
- Aguar 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. 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.
- 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.
- Avila Blas O. J. , Ahumada M. C., Funes H. N., Abril J. C. y Lesino G. (2002). Modelo estadístico estructural de series de velocidad de vientos en la ciudad de Salta. *Energías Renovables y Medio Ambiente*, **10**, 15-22.

ABSTRACT – In this paper a statistical control (testing) is done for a previous statistical structural model that was designed for a daily average speed wind values time series. The model validation is done by introducing an algorithm that includes a dynamic regression model of second order framed in Poisson distribution. This new model allow us to do the statistical analysis for non gaussian time series, which gives it flexibility in order to do forecasting with a high confidence level (95%).

Keywords : wind speed, wind energy, statistic, structural models, forecasting