ASADES

Avances en Energías Renovables y Medio Ambiente
Vol. 15, 2011. Impreso en la Argentina. ISSN 0329-5184

ANALISIS ESTADÍSTICO DE MEDIDAS DE VELOCIDAD DE AIRE

N. Salvo, O. Avila Blas
Universidad Nacional de Salta – Instituto de Energías no Convencionales (INENCO) – Departamento de Física, Facultad de Ciencias Exactas – Avenida Bolivia 5150 - C.P. 4400 – Salta Tel. 54-387-4255424 – Fax 54-387-4255489

E-mail: nahuel@unsa.edu.ar, ojblas@yahoo.com.ar

RESUMEN: En este trabajo se presenta un análisis de medidas realizadas en un túnel de viento. Las medidas fueron realizadas en diferentes puntos, separados espacialmente y en cada lugar de medición se tomaron varios valores para diferentes tiempos. Se plantea a lo largo del trabajo el estudio estadístico de los valores obtenidos con el objeto de generar modelos que permitan realizar pronósticos en el tiempo. Si bien todo el análisis se realiza para valores de velocidad de aire, las mismas herramientas estadísticas pueden ser aplicadas a otras variables ambientales, pensando siempre en una aplicación renovable, por ejemplo datos de radiación solar o velocidad de viento. El análisis de datos en diferentes instantes se lleva a cabo utilizando series de tiempo.
Palabras clave: series temporales, mediciones, predicción, análisis de resultados.

INTRODUCCIÓN

Por lo general cuando se realizan medidas tendientes a caracterizar un determinado prototipo o comprobar una determinada ley o comportamiento de un sistema, éstas se obtienen en zonas previamente identificadas (N. Salvo, *et al*, 2006). El segundo planteo o interrogante surge de pensar en cuántas mediciones hay que realizar y con qué frecuencia de toma de datos, ésto por lo general queda resuelto a partir de llevar a cabo un análisis de los errores experimentales y del sistema de toma de datos.
Suponiendo resuelto lo anterior y que los datos experimentales provienen de un solo lugar, lo que hay que estudiar es qué sucede con los datos adquiridos a partir de una cierta frecuencia de muestreo. Además si el sistema se encuentra en estado estacionario los valores obtenidos no tienen que ser necesariamente iguales o sea que ellos pueden ser diferentes. Por ejemplo si se tiene como sistema a medir un flujo de aire con velocidad media constante y como sensor una punta anemométrica que toma valores de velocidad, con una cierta frecuencia, cada valor adquirido será por lo general diferente al anterior, a pesar de emplear un instrumento muy exacto en su resolución, debido a situaciones aleatorias no controlables. Por lo tanto la diferencia entre valores puede ser motivada por otras causas, como ser algunas características intrínsecas al flujo.
A partir de lo expuesto en el párrafo anterior, el análisis de datos experimentales debe ser realizado empleando herramientas matemáticas no determinísticas o sea algunas que tengan en cuenta aspectos íntimamente relacionados con la probabilidad de ocurrencia de un determinado valor en una serie de datos. Esto respondería, por ejemplo, a la siguiente pregunta, ¿si se volviera a medir una serie de datos de una variable propia del sistema, cuál de los valores de ésta tendría mayor probabilidad de ocurrencia?. También sería factible pensar en que, si bien los datos obtenidos en una medición nos son todos iguales, ellos conservan una cierta relación que pone de manifiesto algún fenómeno que a simple vista no puede ser detectado.
Por lo anterior es importante elaborar algún criterio cuando se analizan datos experimentales a fin de poder explicar el fenómeno bajo estudio de una manera más verosímil. A continuación se demuestra lo anterior a partir del análisis realizado sobre valores de velocidad de aire en un túnel de viento. Si bien el sistema completo se compone de varios puntos de medida sólo se analizan dos por entender que son los más representativos para el análisis. Debe quedar claro que el estudio realizado para dos puntos puede ser extendido a muchos puntos y también es posible generalizar el estudio cuando las condiciones son de no isotropía y estado no estacionario.

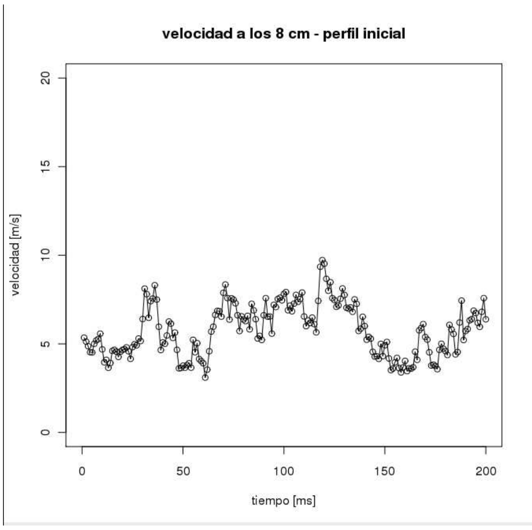
ANÁLISIS DE DATOS

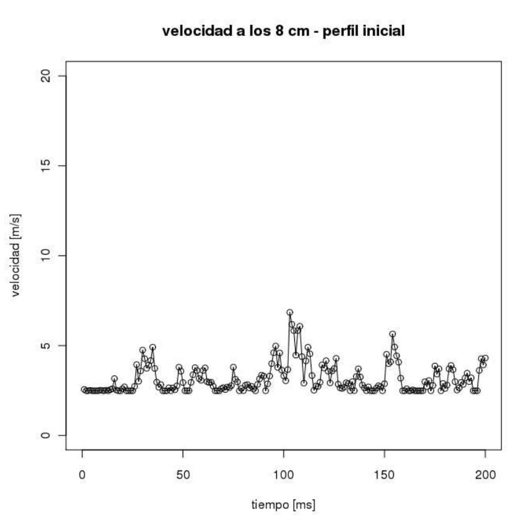
En las figuras que se muestran a continuación (figura 1, a y b) se representan temporalmente 200 valores de velocidad de aire, medidos en un túnel de viento, para dos puntos denotados por “perfil inicial” y “punto5u8cm” y separados espacialmente por una determinada distancia.

Entre cada punto se encuentra un obstáculo. La velocidad del aire en el túnel de viento se mantuvo constante durante toda la experiencia. Los valores fueron adquiridos por una punta anemométrica y en ambos casos se consideró la misma altura (8cm) con respecto a la base del túnel de viento.

Es claro a partir de las figuras 1a y 1b que las condiciones hidrodinámicas de ambos puntos son diferentes, por lo tanto es importante considerar que las variaciones temporales en cada caso responden a una determinada característica del flujo. Esto nos habilita a pensar que ambas mediciones se encuentran correlacionadas o dicho en otras palabras existe cierto vínculo entre ellas.
Por otro lado, los valores presentados de esta forma brindan muy poca información con respecto a la situación que se está analizando, por lo tanto es necesario considerar otros elementos de juicio para explicar mejor el comportamiento de las mediciones.

Una de las explicaciones posibles del por qué las gráficas no corresponden a funciones constantes, sería la existencia de correlaciones temporales entre los valores medidos en cada punto.





(a) ***(b)***

Figura 1: *valores de velocidad de aire en un túnel de viento en dos puntos separados espacialmente. Ambos puntos de medición se sitúan a 8 cm de la base del túnel y se encuentran a lo largo del mismo eje.*

Para analizar los datos utilizando métodos de modelización ARIMA es necesario tener en cuenta las definiciones básicas propias de estos modelos y los siguientes pasos:

- Estudio del comportamiento de la función de autocorrelación temporal.
- Modelado estadístico de las series de tiempo de valores medidos.

ESTUDIO DEL COMPORTAMIENTO DE LA FUNCIÓN DE AUTOCORRELACIÓN TEMPORAL.

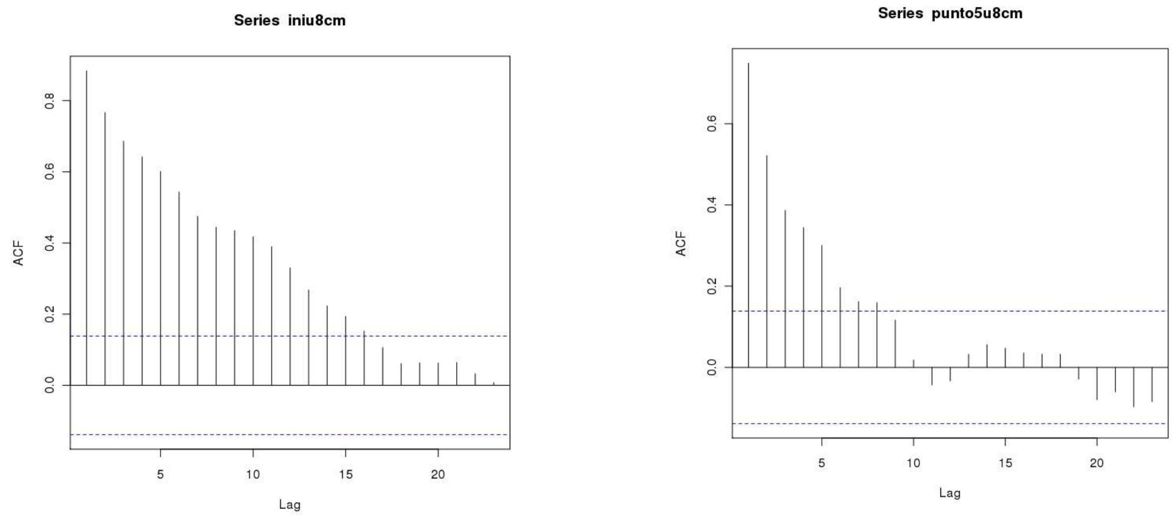
Mediante el estudio de esta función se pueden obtener conclusiones para caracterizar el tipo y grado de vínculo existente entre dos valores medidos en forma consecutiva en el tiempo, o entre más de dos aunque no estén igualmente espaciados. La misma viene definida por (A. C. Harvey, 1996):

$$\hat{\rho}_s = r_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} = \frac{T^{-1}\sum_{j=1}^{T-s}\left(u_{j+s}-\bar{u}\right)\left(u_j-\bar{u}\right)}{T^{-1}\sum_{j=1}^{T}\left(u_j-\bar{u}\right)^2}, \qquad \forall\ 0 \le j < T$$

con ρ_s se representa la función de autocorrelación temporal para un período de longitud s, γ es la función de autocovarianza, el subíndice s representa el período de evaluación de la misma y subíndice 0 es el valor de esta función para un tiempo inicial. *u* representa la variable bajo estudio, en este caso velocidad de aire y T es el número de valores considerados. El estudio del comportamiento de esta función nos habilita a extraer conclusiones a partir de las siguientes consideraciones:

- Se define la banda de confiabilidad como la región limitada por dos rectas paralelas al eje de las abscisas centrada en el origen de coordenadas y de ancho igual al valor $2/\sqrt{T}$, siendo T el número de datos. Cualquier valor de la función que quede comprendido dentro de esta banda, se considera significativamente igual a cero.
- Si p es el primer natural para el que la función es significativamente igual a cero, entonces p se llama el orden del modelo autorregresivo asociado.
- El valor mencionado en el ítem anterior, desde el punto de vista práctico, no siempre nos brinda el modelo óptimo que ajusta a los datos bajo estudio, ya que muchas veces, tomando órdenes menores, se consiguen modelos más simples (menor cantidad de parámetros a estimar). Luego, el valor de p obtenido a partir de la observación de la función de autocorrelación es un valor de orden tentativo. Cuánto más pequeño es el p tomado, al tener que realizar menos estimaciones, se maximiza la propiedad de parsimonia del modelo, ésto es, que el modelo ajusta los datos bajo estudio, con una confiabilidad más alta.

En la figura 2 se muestra la función de autocorrelación de los datos para cada punto de medida. Se destaca que en ambos casos, las bandas de confiabilidad (lineas de a trazos horizontales) tienen el mismo ancho, dado que se ha tomado el mismo número de mediciones. En la gráfica a) puede verse un decaimiento muy lento de la función y la misma toma un valor cuasi nulo a partir del valor que ocupa la posición 16. Luego de ésta, la función tiene un decaimiento más abrupto, sin salirse de la banda.
En la gráfica b) se observa un decrecimiento mucho más pronunciado que en el caso anterior, siendo el primer valor de cuasi nulidad, el que ocupa la posición 8. Se destaca además que luego de este valor, la función si bien se mantiene dentro de la banda de confiabilidad, manifiesta un comportamiento de tipo oscilatorio, lo que haría pensar en la presencia de un posible ciclo oculto, pero de una amplitud muy pequeña.



(a) *(b)*

Figura 2: *función de autocorrelación de datos para cada punto de medida.*

La forma general de un modelo autorregresivo de orden p, para una serie de tiempo de la forma $\{u_t\}$, denotado con la simbología AR(p), está expresado por la siguiente ecuación:

$$u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \cdots + \alpha_p u_{t-p} = u_{t+1}$$

donde los coeficientes α son las ponderaciones o pesos de la variable en los diferentes instantes de tiempo t. En muchas situaciones, el estudio de la función de autocorrelación debe ser complementado con el de la denominada función de autocorrelación parcial. La misma se define por la siguiente expresión:

$$\rho_{1a} = Correl(u_2, u_1) = \rho_1$$

$$\rho_{sa} = Correl\left(u_{s+1} - P_{\bar{s}p}\{1, u_2, \cdots, u_s\} u_{s+1} ; u_1 P_{\bar{s}p}\{1, u_2, \cdots, u_s\} u_1 \right) \quad para \quad s \geq 2$$

Se representa con “Correl” la función de autocorrelación entre dos valores de la variable *u*, los cuales no necesariamente tienen que ser consecutivos en el tiempo. Su cálculo no es tan directo como en el caso de la función anterior, pero todos los software específicos usados para el modelado de series de tiempo realizan el cálculo en forma discreta de ella y permiten su representación gráfica.

El estudio del comportamiento de esta función habilita a extraer conclusiones a partir de las siguientes consideraciones:

- Se define la banda de confiabilidad como la región limitada por dos rectas paralelas al eje de las abscisas centrada en el origen de coordenadas y de ancho igual al valor $2/\sqrt{T}$, siendo T el número de datos. Cualquier valor de la función que quede comprendido dentro de esta banda, se considera significativamente igual a cero.

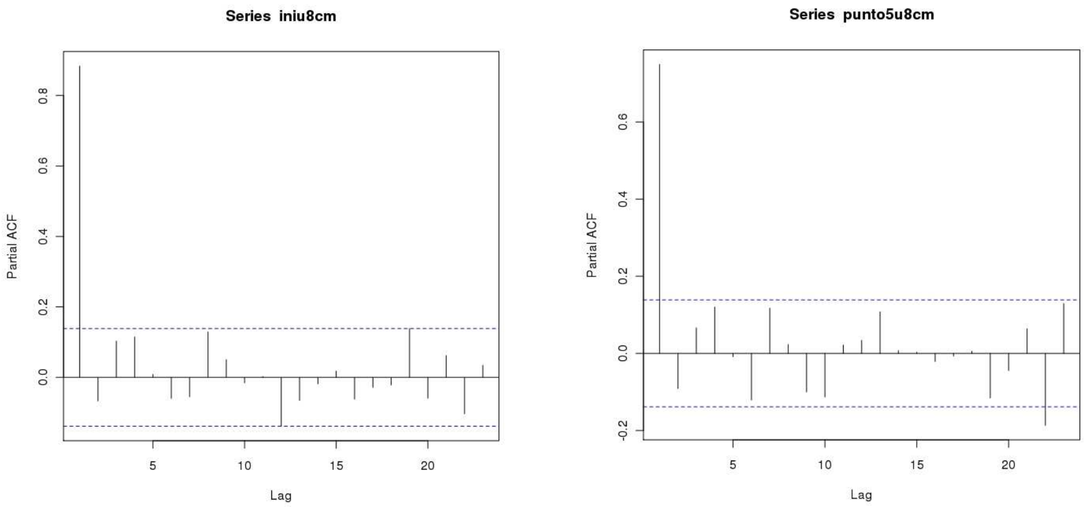
- Si q es el primer natural para el que la función es significativamente igual a cero, entonces q se llama el orden del modelo de promedios móviles asociado.

- El valor mencionado en el ítem anterior, desde el punto de vista práctico, no siempre nos brinda el modelo óptimo que ajusta a los datos bajo estudio, ya que muchas veces, tomando órdenes mayores, se consiguen modelos más robustos (aunque la cantidad de parámetros a estimar aumente ligeramente). Luego, el valor de q obtenido a partir de la observación de la función de autocorrelación parcial es un valor de orden tentativo. El procedimiento general para garantizar la obtención del modelo óptimo (modelo que ajusta los datos bajo estudio, con una confiabilidad más alta), consiste en general, en tomar aquel orden q que tenga la propiedad de que la función de autocorrelación parcial decrezca rápidamente a cero, a medida que aumentamos el número de períodos de observación.

Para los datos medidos en los dos puntos bajo estudio, esta función se muestra en la figura 3. En la misma se puede observar que si pretendemos ajustar nuestros datos con modelos de promedios móviles, los órdenes tentativos deben ser mayores o iguales a 2, dado que en general el segundo valor de la función en cuestión es significativamente igual a cero.
Dada una serie de tiempo de la forma $\{\varepsilon_t\}$, la forma general de un modelo de promedios móviles de orden q, denotado por con la simbología MA(q), está expresado por la siguiente ecuación:

$$u_t = \varepsilon_t + \beta_q \varepsilon_{t-q} + \cdots + \beta_2 \varepsilon_{t-2} + \beta_1 \varepsilon_{t-1}$$

Los coeficientes β son los pesos o ponderaciones asociados a los diferentes valores de la serie.



(a) *(b)*

Figura 3: *función de autocorrelación parcial de datos para cada punto de medida.*

Si bien, en muchas situaciones prácticas, es suficiente tomar modelos autorregresivos ó de promedios móviles para poder generar sintéticamente valores de la variable bajo estudio, en otros casos ésto no resulta un procedimiento confiable. Ésta es la situación que se adecúa al tratamiento de los datos en el presente trabajo.
Cuando no es factible el uso de una sola forma de modelado, resulta conveniente plantear el uso de un modelo que sea una combinación de los anteriores. Tal formulación integrada se conoce por modelos ARIMA(p,d,q) donde p es el orden del modelo autorregresivo y q es el orden del modelo de promedios móviles correspondiente. El valor d se conoce como el orden de diferenciación y es usual para obtener modelos confiables operar con los datos de la serie realizando diferencias entre valores, con un lapso intermedio definido por d. En general un modelo de este tipo se expresa de la siguiente forma:

se define el operador de diferenciación o de rezagos L, como: $L(u_t) = u_{t-1} \wedge L^d(u_t) = u_{t-d}$ con d = orden de diferenciación

En base a este operador, un modelo ARIMA(p,d,q) para la serie $\{u_t\}$ se construye haciendo: $(1-L)^d u_t$

MODELADO ESTADÍSTICO DE LAS SERIES DE TIEMPO DE VALORES MEDIDOS.

Para los datos considerados y en base al comportamiento de las funciones de autocorrelación total y parcial, se procedió a seleccionar los parámetros p, d, y q que resultaban tentativos para el modelado de los valores medidos. Las ternas de parámetros que resultaron apropiadas son: (1,1,3), (1,1,4) para los valores del perfil inicial; y (1,1,1), (1,1,2) para los del segundo punto considerado.

PERFIL INICIAL

Altura 8 cm	datos	ar1	ma1	ma2	ma3	ma4	sigma^2	log likelihood	aic
arima(x = iniu8cm, order = c(1, 1, 3))	200	0,2387	-0,2779	-0,1591	-0,1114				
	s.e.	0,3908	0,3874	0,0788	0,1024		0,461	-205,41	418,82
arima(x = iniu8cm, order = c(1, 1, 4))	200	-0,122	0,084	-0,1668	-0,1759	-0,0739			
	s.e.	0,6489	0,3454	0,077	0,1266	0,1305	0,4607	-205,33	420,67

PUNTO5U8CM

Altura 8 cm	datos	ar1	ma1	ma2	sigma^2	log likelihood	aic
arima(x = punto5u8cm, order = c(1, 1, 1))	200	0,7635	-1,0000				
	s.e.	0,0473	0,0376		0,2977	-163,45	330,91
arima(x = punto5u8cm, order = c(1, 1, 2))	200	0,7055	-0,8626	-0,1325			
	s.e.	0,0750	0,1185	0,0945	0,2957	-162,54	331,08

Tabla 1: Parámetros de los modelos ARIMA

La tabla 1 se interpreta de la siguiente manera: en la columna ar1, ma1, ma2, etc. figuran los coeficientes estimados correspondientes a la componente autorregresiva o de promedios móviles del orden indicado por el número que forma parte del nombre de la columna. En la casilla s.e., se muestra el desvío estándar correspondiente a la estimación del parámetro según lo indicado en el párrafo anterior. En las últimas tres columnas figuran los valores de la varianza total del modelo (sigma^2), el logaritmo de la función de verosimilitud (log likelihood) y el indicador de testeo de Akaike (aic). (A. C. Harvey, 1996)

A partir de los valores incluidos en la respectivas tablas se obtienen los siguientes resultados:

- Para el perfil inicial, es conveniente tomar un modelo ARIMA(1,1,3) puesto que tiene asociado el mayor valor en módulo, del logaritmo de la función de verosimilitud (205,31) y además contiene en menor número de parámetros a estimar.
- Para el punto5u8cm, se aconseja tomar un modelo ARIMA(1,1,1) ya que tiene el menor número de parámetros a estimar, que resulta ser dos, y una función de verosimilitud en logaritmo máxima en módulo (163,45).

Con esta selección de modelos se procedió a realizar el proceso de generación de valores sintéticos o valores pronosticados, con una confiabilidad muy alta (95%), tal como se muestra en las figuras 4 y 5.

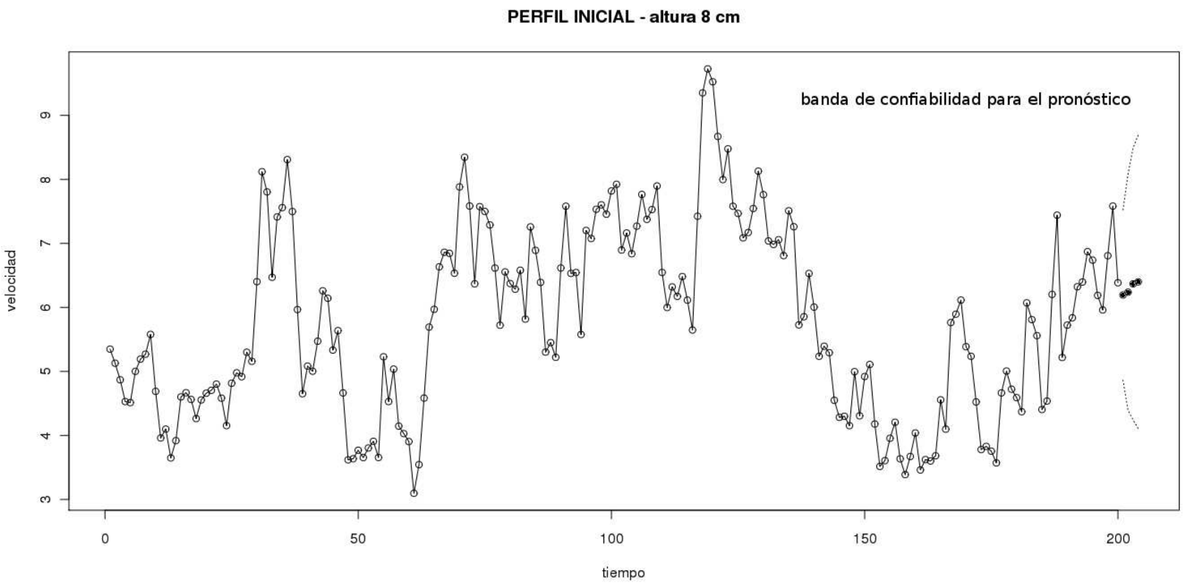


Figura 4: *Valores generados para el modelo ARIMA(1,1,3)*

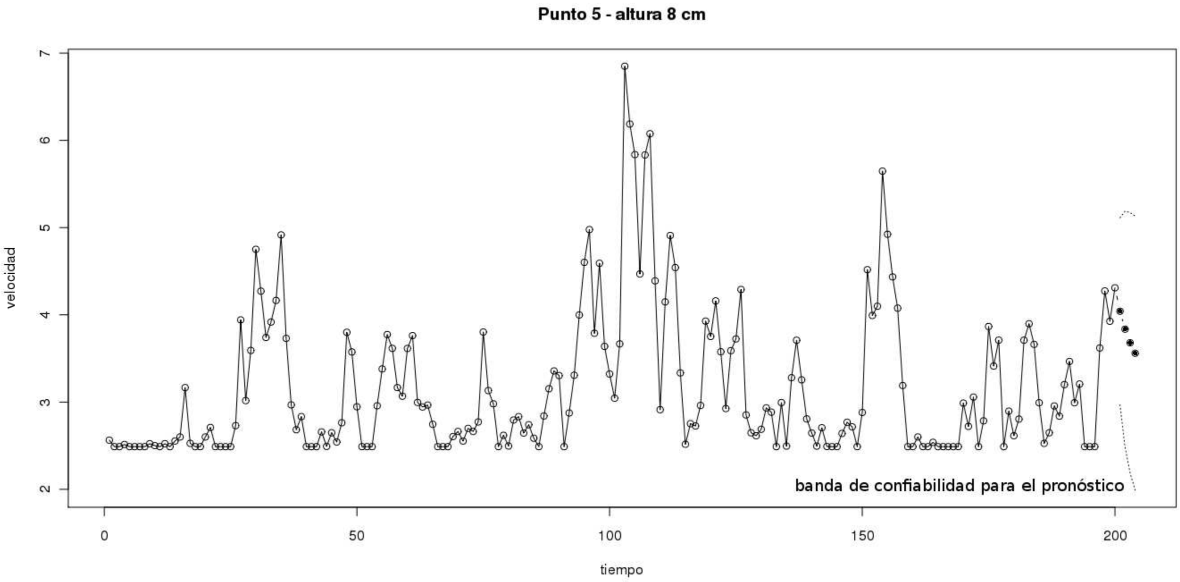


Figura 5: *Valores generados para el modelo ARIMA(1,1,1)*

En ambos gráficos, se pronostican cuatro valores hacia adelante en el tiempo, representados por los puntos negros, los cuales quedan incluidos dentro de las respectivas bandas de confiabilidad asociadas al proceso e indicadas por curvas de puntos, que son de un ancho variable.

Si bien se pronosticaron cuatro valores en el tiempo, se puede extender el número de los mismos teniendo en cuenta que este proceso debe ser controlado de alguna manera, por ejemplo tratando que la banda de confiabilidad no cambie significativamente en su ancho.

En el caso del perfil inicial, es importante destacar que los valores medidos se mantienen correlacionados más fuertemente en el tiempo ya que la función de autocorrelación tiene un decaimiento suave. El proceso ARIMA que los describe es de la forma (1,1,3), lo cual significa tener que estimar tres coeficientes para la componente de promedios móviles a fin de garantizar que el proceso de pronóstico se realice sin perder el nivel de confiabilidad.

Para el otro punto de medida, se necesita estimar un menor número de parámetros puesto que la función de autocorrelación presenta un decaimiento más rápido.

Uno de los aspectos observados al inicio de este trabajo fue el de pensar que las mediciones en los dos puntos de toma de datos, se encontraban correlacionados o que existía cierto vínculo entre ellos. La respuesta a esta conjetura viene indicada por la propiedad de que en ambos modelos finales, el orden de diferenciación es d=1. Esto quiere decir que a una misma altura de toma de datos (propiedad espacial) a los fines de realizar un proceso de modelado estadístico, es necesario tomar la serie original y realizarle una diferenciación de orden d = 1, en ambos puntos.

La metodología presentada se encuentra programada con el software estadísto R, que es de distribución libre. (R. Gentleman, *et al*; R. H. Shumway *et al*, 2006; J. D. Cryer *et al*, 2008)

CONCLUSIONES

Se ha desarrollado una metodología de trabajo que permite pronosticar valores a partir de un modelo estadístico de series de tiempo (ARIMA) para mediciones realizadas en un túnel de viento, con una muy alta confiabilidad (95%). Este tipo de procedimiento es también aplicable para generar temporalmente valores sintéticos de otras variables.

Mediante la determinación del orden del modelo ARIMA, se pudo detectar la existencia de una correlación espacial entre los dos puntos de mediciones.

BIBLIOGRAFIA

1 – N. Salvo, I. De Paul, “Análisis fluidodinámico para una vivienda tipo, considerando distintas orientacione. Mediciones en un túnel de viento” - ERMA, Vol 18 (2006) pp 75 – 83.

2 – R. Gentleman, R. Ihaka, “The R project for statistical computing” - software libre resultado de la implementación GNU.

3 – R. H. Shumway, D. S. Stoffer, “Times Series Analysis and Its Applications” Second Edition , (2006) Springer Verlang - ISBN-10: 0-387-29317-5 - ISBN-13: 978-0387-29317-2

4 – J. D. Cryer, K.-Sik Chan, “Time Series Analysis With Applications in R ” - Second Edition, (2008) Springer Verlang -

ISBN: 978-0-387-75958-6 e-ISBN: 978-0-387-75959-3

5 – A. C. Harvey, “Forecasting, structural time series models and the Kalman filter” - (1996) Canbridge University Press – ISBN: 0-521-32196-4

ABSTRACT

In this paper an analysis of measurements taken in a wind tunnel is presented. Measurements were taken at different points, spatially separated and in each measurement site several values were taken at different times. Along the work an statistical analysis of the obtained values is set in order to generate modells that allow us to do the timing forecasting. Despite of the hole analysis is performed for values of air velocity, the same statistical tools might be applied to other environmental variables, always thinking about on a renewable application, for example solar radiation or wind velocity. The data analysis at different times is carried out using time series.

Keywords: time series, measurements, forecasting, results analysis