

A stochastic model of wind velocity in Lake Maracaibo

Salvador Pintos

*Instituto de Cálculo Aplicado (ICA), Facultad de Ingeniería
Universidad del Zulia, Apartado 526. Maracaibo, Venezuela*

Abstract

In this paper hourly records of wind velocity were analyzed as a Multivariate Time Series. Data were collected from three meteorological stations on the coast of Lake Maracaibo. The fundamental result was the identification with a multivariate ARMA model that was estimated and validated for every month of the year. Before the identification, the seasonal component was eliminated by means of a suitable transformation. The result of this investigation is useful for contingency plans in the area.

Key words: Wind, velocity, multivariate ARMA.

Modelo estocástico de la velocidad del viento en el Lago de Maracaibo

Resumen

Se analiza aquí registros horarios de la velocidad del viento en tres Estaciones Meteorológicas ubicadas sobre las costas del Lago de Maracaibo, como una Serie de Tiempo Vectorial. El resultado fundamental es su identificación, previa eliminación del componente estacional diario, con un modelo ARMA Vectorial que se estima y valida para cada mes del año. Este estudio será de utilidad para planes de contingencia en el área.

Palabras claves: Velocidad, viento, modelos ARMA multivariados.

Introducción

Modelar la velocidad horaria del viento en el Lago de Maracaibo a corto plazo es útil para diferentes actividades productivas. En especial en la Industria Petrolera es decisivo para pronosticar la evolución de un derrame de petróleo [1] y [2]. La peculiar circulación ciclónica antihoraria de la velocidad del viento [2] hace necesario un estudio temporal y espacial de la misma. En la secuencia horaria de la velocidad del viento existe una fuerte correlación entre observaciones consecutivas. Es por ello que un modelo que simule la velocidad debe usar esta información secuencial. Los estudios disponibles sobre vientos en la región del Lago de Maracaibo [3-6] y [2], así como los resúmenes del MARNR y las F.A.V. no preservan la correlación secuencial. El obje-

tivo de este trabajo es modelar la velocidad del viento en el Lago con el propósito de pronosticar la velocidad en el corto plazo. Para ello se estudian los datos de tres Estaciones Meteorológicas (M.A.R.N.R.) Santa Bárbara, CANAL VOC y Barinitas ubicadas sobre las costas del Lago al Sur, Este, y Oeste de la Bolsa, (zona ésta de explotación petrolera) ante la falta de información aguas adentro. Diferentes modelos estocásticos de la velocidad del viento han sido formulados, que incluyen series de tiempo, cadenas de Markov [7], etc. Aquí, para cada estación se modela la velocidad del viento como una Serie de Tiempo Vectorial. Se ejemplifica con datos de Santa Bárbara aunque el modelo hallado es común a las tres estaciones, difiriendo sólo en los parámetros estimados que reflejan las características de la zona.

En la primera parte de este trabajo se trata la eliminación del ciclo horario para transformar el proceso en estacionario y se estiman los parámetros de la regresión. En la segunda, se establece una transformación para obtener distribuciones normales en las componentes del proceso estacionario. En la tercera, se formula el modelo VAR(1) y se obtienen los estimadores de los parámetros. En la cuarta, se hace el diagnóstico del modelo. En la quinta, se estudian las características del pronóstico para este modelo y sus resultados.

Estacionalidad, Ciclo Diario

La variación estacional de los vientos [2], con marcada influencia tanto en la dirección como en la velocidad, justifica el desglose de los datos en doce estaciones, una para cada mes del año. Para cada mes se analizó los datos de los años disponibles (1978-1986). Los datos (incompletos) contienen los registros horarios originales del M.A.R.N.R. (1978-1986) de la celeridad en km/h y la dirección en una escala discreta de 16 divisiones cada una a 22.5 grados de las contiguas, con el 0 en el Norte, 90 en el Este, etc. Se consideró más apropiado usar las componentes ortogonales de la velocidad tomando como variables:

$$V_x = \text{Celeridad} \cdot \cos(\alpha)$$

$$V_y = \text{Celeridad} \cdot \sin(\alpha) \quad (1)$$

donde α es el ángulo correspondiente a la dirección.

Para un mes cualquiera V_x y V_y están altamente correlacionadas y ambas muestran la influencia de la hora del día observándose un comportamiento estacional. Sin embargo, la correlación parcial entre ambas al eliminar el efecto de la hora se reduce considerablemente.

Las Figuras 1 y 2 muestran el periodograma de V_x y V_y para el mes de enero del 79 (S. Bárbara) obtenido con el procedimiento SPECTRA del SAS. Obsérvese la predominancia en el espectro de las ondas diurnas y semidiurnas período 24 y 12. Periodogramas similares se observan para cualquier otro mes del año. Se han propuesto distintas metodologías para la eliminación de ciclos en climatología, [8]. Con el objeto de lograr que el proceso sea estacionario se usaron dos metodologías para eliminar la estacionalidad horaria:

- A. 24 variables ficticias cada una de ellas asociada a la media de la hora respectiva.
- B. Eliminación de las ondas asociadas a las frecuencias dominantes en el periodograma.

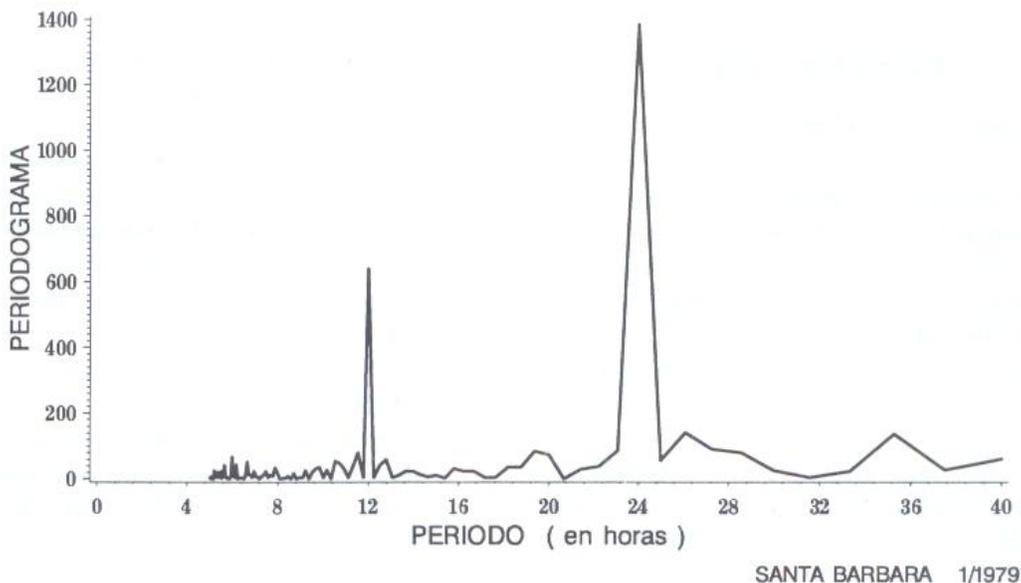


Figura 1. Análisis espectral de V_x .

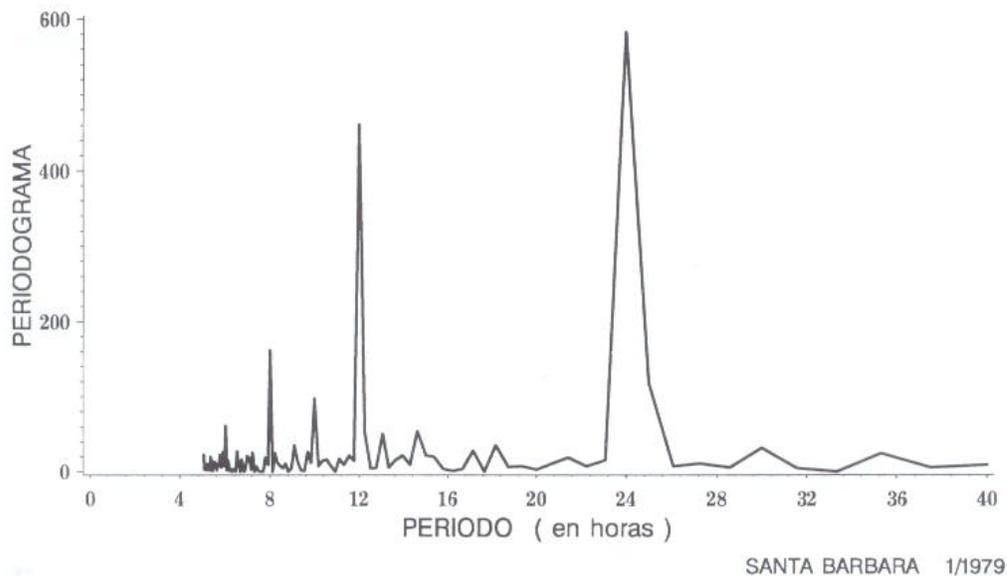


Figura 2. Análisis espectral de Vy.

Se determinaron por regresión los estimadores de los parámetros de ambos modelos, obteniéndose en el caso A un mejor coeficiente de determinación R^2 que en el caso B, pero no suficiente como para justificar la introducción de 24 parámetros en el modelo en vez de 4. Por otra parte los resultados de las pruebas t del método A no son significativos para la mayoría de los parámetros al nivel de riesgo del 5%, en cambio los asociados a las ondas diurnas y semidiurnas tienen un valor de probabilidad en torno al 0.1%.

Para el caso B, más parsimonioso, que es el que se siguió, para cada variable se formula el modelo respectivo:

$$\begin{aligned}
 V_x &= I_x + a \cos\left(\frac{2\pi t}{24}\right) + b \sin\left(\frac{2\pi t}{24}\right) + \\
 &c \cos\left(\frac{2\pi t}{12}\right) + d \sin\left(\frac{2\pi t}{12}\right) + W_x \\
 V_y &= I_y + e \cos\left(\frac{2\pi t}{24}\right) + f \sin\left(\frac{2\pi t}{24}\right) + \\
 &g \cos\left(\frac{2\pi t}{12}\right) + h \sin\left(\frac{2\pi t}{12}\right) + W_y
 \end{aligned}
 \tag{2}$$

Estimados las constantes I_x e I_y y los parámetros a, b, c, d, e, f, g, h el vector de componentes W_x y W_y se comporta como un proceso estacionario con medias 0. Este proceso de eliminación de componentes estacionales reduce además la desviación standard en un 30% aproximadamente. Las variables W_x y W_y tienen una correlación menor que la de las variables V_x y V_y originales pero no soportan la prueba de independencia, de modo que es necesario modelarlas conjuntamente. Para los datos del mes 1/79, por ejemplo, los valores de probabilidad asociados son: 0.01 para V_x, V_y y 0.04 para W_x, W_y .

Transformación Normal

Las variables W_x y W_y no soportan la prueba de normalidad, de modo que se estudió para todos los períodos y para todos los años una transformación de la forma:

$$\begin{aligned}
 X &= \text{signo}(W_x) | W_x |^\lambda \\
 Y &= \text{signo}(W_y) | W_y |^\lambda
 \end{aligned}
 \tag{3}$$

observándose mediante una modificación de la prueba de Kolmogorov-Smirnov que para $\lambda=0.8$ o $\lambda=0.85$ la transformación indicada hace que las variables X e Y no rechacen la prueba de

normalidad, lo cual es necesario para la determinación de los intervalos de confianza. Se obtienen mejores resultados para $\lambda = 0.8$, valor que se usará en lo que sigue.

Modelo Auto-Regresivo

Un análisis separado para X e Y con la metodología de Box-Jenkins permite observar en las funciones de autocorrelación muestral, total y parcial, las características de un proceso estacionario e identifican un modelo AR(1), pero no pueden modelarse separadamente debido a la correlación entre ambas variables. Se sugiere entonces un modelo Vectorial AutoRegresivo de orden 1, VAR(1). Sea Z el vector columna de componentes X e Y , el modelo se formula:

$$Z_t = A Z_{t-1} + B E_t \quad (4)$$

Donde A y B son matrices 2×2 y E el vector de ruido blanco donde los E_t son independientes dos a dos, con valor esperado 0 y varianza 1. Se supone además, que el ruido para el tiempo t es independiente del valor de la serie Z en el tiempo $t-1$. La estimación de A y B a partir de las ecuaciones de Yule-Walker teóricas [3] y [9] consiste en identificar los momentos muestrales con los teóricos y deducir así los parámetros matriciales.

Para ello se definen las matrices de covarianza C_0 y C_1 teóricas del vector Z :

$$\begin{aligned} C_0 &= E[Z_t Z_t^T] \\ C_1 &= E[Z_t Z_{t-1}^T] \end{aligned} \quad (5)$$

donde E es el valor esperado y ' T ' significa el operador traspuesta de una matriz.

Postmultiplicando primero por Z_t^T en (4) y luego por Z_{t-1}^T en (4), haciendo uso de (5), y tomando valores esperados en ambas ecuaciones teniendo en cuenta los supuestos del modelo se obtiene:

$$\begin{aligned} A &= C_1 IC ; IC = INV(C_0) \\ BB^T &= C_0 - C_1 IC C_1^T \end{aligned} \quad (6)$$

La determinación de A es única, en cambio B admite distintas soluciones; se optó por considerar a B como una matriz triangular subdiagonal que conduce a una solución única dado que BB^T es definida positiva puesto que es la covarianza condicional de Z_t dada la observación pasada Z_{t-1} [9].

Los estimados de A y B a partir de los valores muestrales de C_0 y C_1 se determinaron para cada mes y promediados con sus homólogos de cada año. Los estimados de A y B se muestran estables para idéntico período (mes). Las matrices siguientes, por ejemplo, corresponden a las estimaciones de A para el mes de octubre en los años 78, 79, 80 y 81:

$$\begin{array}{cc} 78 & 79 \\ \left[\begin{array}{cc} 0.586 & -0.024 \\ 0.059 & 0.429 \end{array} \right] & \left[\begin{array}{cc} 0.534 & -0.050 \\ 0.074 & 0.470 \end{array} \right] \\ 80 & 81 \\ \left[\begin{array}{cc} 0.545 & -0.004 \\ 0.067 & 0.432 \end{array} \right] & \left[\begin{array}{cc} 0.612 & 0.003 \\ 0.096 & 0.534 \end{array} \right] \end{array}$$

Obsérvese que los valores significativos de la diagonal principal representan la asociación entre las mismas componentes para horas sucesivas y los de la diagonal secundaria, el efecto cruzado de las componentes.

En cuanto a las estimaciones respectivas de B :

$$\begin{array}{cc} 78 & 79 \\ \left[\begin{array}{cc} 1.789 & 0. \\ 0.398 & 1.59 \end{array} \right] & \left[\begin{array}{cc} 1.700 & 0. \\ 0.266 & 1.568 \end{array} \right] \\ 80 & 81 \\ \left[\begin{array}{cc} 1.835 & 0. \\ 0.136 & 1.630 \end{array} \right] & \left[\begin{array}{cc} 1.586 & 0. \\ -0.006 & 1.283 \end{array} \right] \end{array}$$

nótese que las varianzas del vector residual $B E_t$ se mantienen estables ya que la influencia de $B_{2,1}$ en BB^T es poco significativa.

Diagnóstico

El diagnóstico del modelo se hizo aplicando distintas pruebas a los residuales, obteniéndose:

- a) Independencia de los residuales.
- b) Las funciones de autocorrelación muestral, total (FACM) y parcial (FACPM), de los residuales soportan la hipótesis de existencia de ruido blanco [3], como se observa en las Figuras 3 y 4 para los datos del residual de X correspondiente al mes 1/79.
- c) Las correlaciones agrupadas mediante la prueba de Box-Ljung no contradicen la hipótesis de existencia de ruido blanco [3].
- d) El análisis espectral de los residuales satisface la prueba del cociente entre la máxima amplitud y la media de las amplitudes (prueba de Fischer).
- e) El espectro integrado soporta la prueba de Kolmogorov-Smirnov al nivel de significación del 5%, no desviándose de la hipótesis nula de uniformidad [10], como se observa en las Figuras 5 y 6 para los datos de los residuales de X e Y correspondientes al mes 1/79.

Pronóstico

Si se define el pronóstico del vector Z para L unidades de tiempo hacia adelante $Z_t(L)$ como la esperanza condicional:

$$Z_t(L) = E [Z_{t+L} | Z_k, k = 1, 2, \dots, t] \tag{7}$$

Resulta que para el modelo VAR(1) se cumple:

$$Z_t(L) = A^L Z_t \tag{8}$$

Para los estimados de A obtenidos en todos los meses y todos los años, los coeficientes de la diagonal principal oscilan entre 0.4 y 0.7, indicando una correlación positiva pero no muy fuerte entre observaciones consecutivas, por lo cual las potencias de A tienden rápidamente a 0 haciendo que el pronóstico a largo plazo (24 horas) coincida estrictamente con la componente determinística estacional. Sin embargo, es en las primeras horas de un pronóstico, $Z_t(L)$, donde se observa las ventajas de este modelo al aprovechar la estructura autoregresiva del mismo, mejorando la predicción determinística estacional.

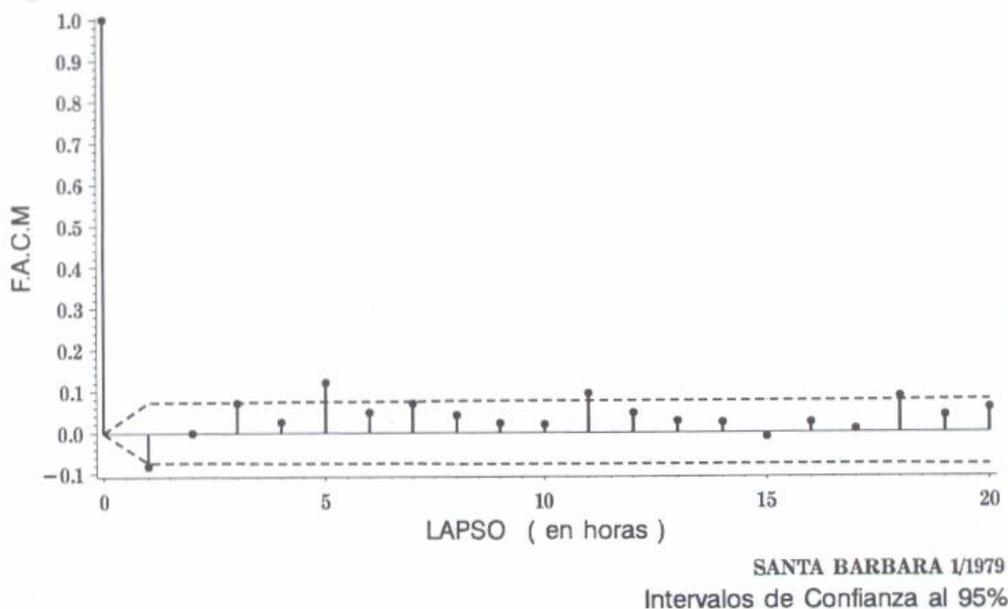


Figura 3. FAC del residual de X.

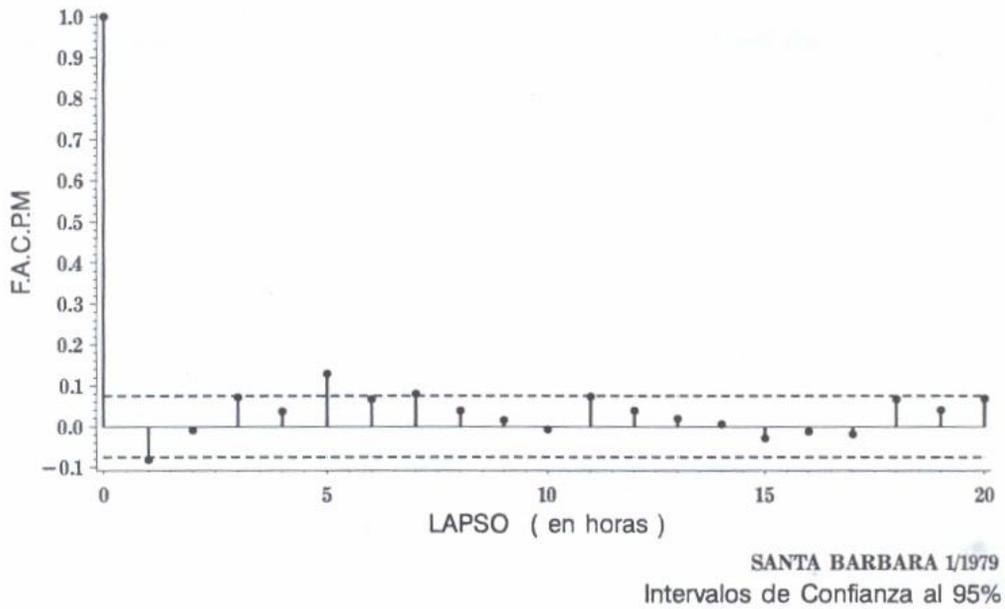


Figura 4. FACP del residual de X.

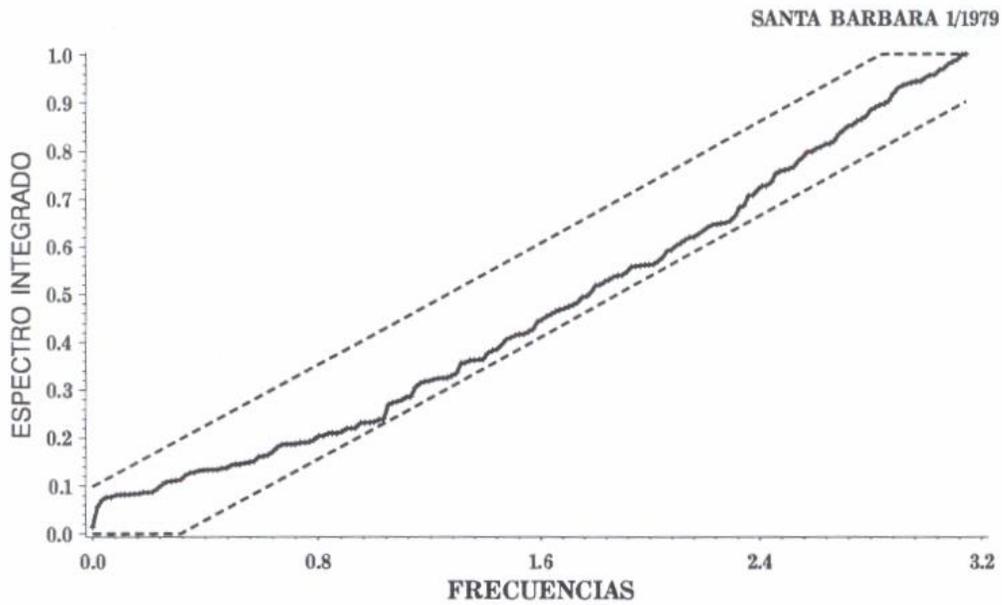


Figura 5. Test del residual de X. Intervalos de Confianza al 95%.

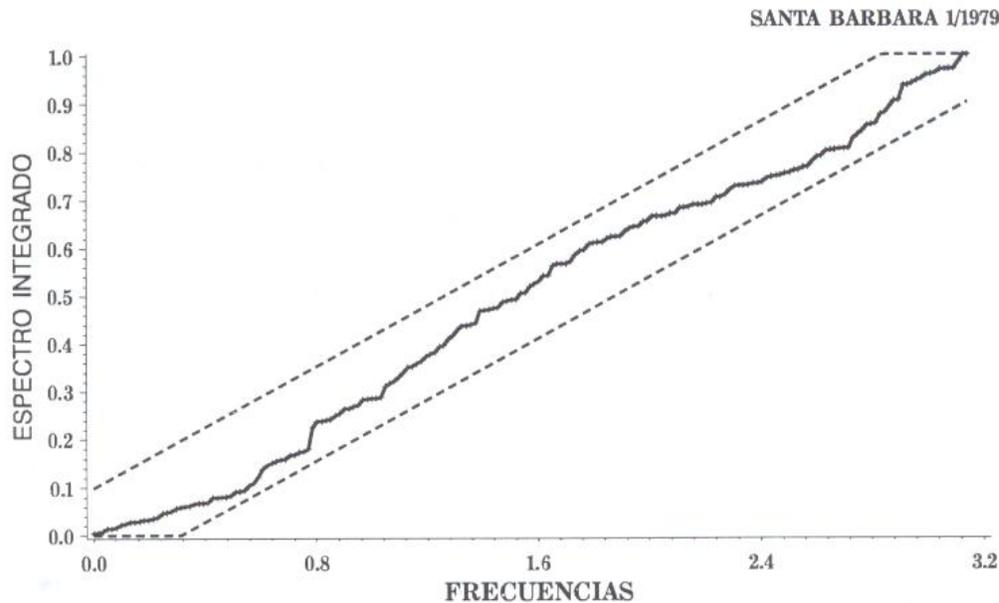


Figura 6. Test del residual de Y. Intervalos de Confianza al 95%.

Y son éstas las horas claves para predecir por ejemplo la evolución de un derrame de petróleo.

Aplicando el modelo a meses no incluidos en la obtención del mismo se obtiene un razonable ajuste en el corto plazo. La simulación aleatoria de datos con el modelo produce observaciones con características similares a la de los datos reales. Para optimizar la eficiencia del modelo se recomienda aplicarlo junto a un sistema de adquisición de datos en tiempo real.

Referencias Bibliográficas

1. Octavio, K.: "Modelo de la trayectoria y comportamiento de derrames de petróleo" RIPPET, INTEVEP 1986.
2. Parra Pardi y asociados: Estudio para la determinación de áreas sensibles en el sistema del Lago de Maracaibo. 1986.
3. Box, G.; Jenkins, G.: Time Series Analysis: Forecasting and Control. Holden Day Inc. 1976.
4. De Oteiza, P.: La dirección del viento en Maracaibo, LUZ, Facultad de Arquitectura, 1983.
5. De Oteiza, P.: La velocidad del viento en Maracaibo, LUZ, Facultad de Arquitectura, 1983.
6. Masciangioli, P.: "Patrones de vientos y corrientes en Puerto Miranda", RIPPET, INTEVEP 1987.
7. Roldan, J.; Garcia, A.; Losada, A.: A Stochastic Model for Wind Occurrence. J. Appl. Meteor., Vol. 21, 1982, 740-744.
8. Fuenzalida, H.; Rosenblut, B.: Distortion Effects of the Anomaly Method of Removing Seasonal or Diurnal Variations from Climatological Time Series J. Clim. Appl. Meteor. Vol. 25 1986, 728-731.
9. Rodriguez Iturbe, I.; Bras, R.: Random Functions and Hydrology, Addison Wesley, Massachusetts, 1985.
10. Jenkins, G.; Watts, D.: Spectral Analysis and its applications. Holden Day Inc. 1968, 234-236.

Recibido el 8 de Octubre de 1993

En forma revisada el 15 de Noviembre de 1994