

**Caracterización y análisis de series de tiempo de enfermedades respiratorias agudas en la provincia de Camagüey.**

**Characterization and analysis of acute respiratory diseases time series from Camagüey.**

**Autores:** Betancourt JA<sup>1</sup>, Brito HS<sup>2</sup>, Ortiz EH<sup>4</sup> 3 y Rodríguez NS<sup>5</sup>

1. Miembro del Centro de Medicina y Complejidad (CMC) del Instituto Superior de Ciencias Médicas. Máster dedicado a la gestión de proyectos biomédicos y sistemas complejos. Correo: [josebetancourt@infomed.sld.cu](mailto:josebetancourt@infomed.sld.cu) teléfono 288668
2. Lic. en Matemática del CMC
3. Lic. En Física del CMC
4. Lic. en Física de la Unidad provincial de Análisis y Tendencias del Centro Provincial de Higiene y Epidemiología.

## **Resumen**

**Introducción.** En Cuba existe un excelente sistema de vigilancia, pero tiene que enfrentar el reto de constantes cambios demográficos y movimientos de poblaciones dentro y fuera del país, entre otros. Existe un eficiente sistema de alerta y respuesta contra la presentación de enfermedades, pero estos programas deben tener en cuenta el presupuesto para las actividades de la salud pública y la complejidad de la propagación; el uso de modelos matemáticos podría apoyar este esfuerzo.

El objetivo de este trabajo es caracterizar una serie de tiempo de enfermedades respiratorias agudas en la provincia de Camagüey y evaluarla con un método adecuado a su comportamiento.

**Métodos.** Se utilizaron tasas de la serie de tiempo semanal de enfermedad respiratoria aguda reportada entre los años 2000 al 2007 de la provincia de Camagüey, Cuba... Para caracterizar estas series se utilizó el programa Matlab V-7,4-2006. La estacionariedad se evaluó con el método de auto correlación simple (FSA), el comportamiento de linealidad se realizó con la prueba BDS, con el método basado en la geometría espectral se determinó si la serie era estocástica o determinista. . Se comparan los errores en las predicciones con los modelos ARMA y GARCH. y se analizan los errores y las curvas de predicción.

**Resultados.** Se demuestra el comportamiento no estacionario de la serie, basado en la Función de auto correlación simple (FSA). La prueba BDS indica que la serie estudiada es no lineal. La prueba de Geometría espectral concluye que la serie proviene de un sistema estocástico. En esta serie estacionaria, no lineal y estocástica resultó más adecuado el modelo GARCH que el ARMA, dado el menor error y más ajuste de la curva de predicción.

**Discusión.** Los resultados obtenidos en este trabajo demuestran la necesidad de clasificar las series de tiempo antes de ser analizadas, para proponer el modelo más adecuado.

**Conclusión.** La serie IRA estudiada resultó ser (a) no lineal, (b) no estacionaria y (c) estocástica y la predicción fue superior en este caso, con el modelo GARCH comparado con el ARMA.

Palabras claves: Predicción epidemiológica, series de tiempo epidemiológica, caracterización series de tiempo, modelo GARCH, epidemiología sistemas complejos.

## **Summary.**

**Introduction.** There is in Cuba an excellent epidemiological surveillance system and early warning, but the challenge it is to face national and international population movement .All these programs are expensive. The use of modern mathematical model might help to reduce cost and to save time.

The objective of this work is to characterize a time series of acute respiratory disease in the county of Camagüey and to evaluate it with an appropriate method according to its behavior.

**Methods.** The rates weekly time series of the series of acute respiratory disease reported among the years 2000 at the 2007 of the county of Camagüey were used, Cuba... To characterize these series the program Matlab V-7,4-2006 it was used. The estacionarity was evaluated with the method of simple auto correlation (FSA), the linearity behavior was carried out with the test BDS, with the method based on the spectral geometry it was determined if the series was stochastic or deterministic. The errors are compared in the predictions with the models ARMA and GARCH..

**Results.** The non stationary behavior of the series is demonstrated, based on the Function of simple auto correlation (FSA). The test BDS indicates that the studied series is nonlinear. The test of Spectral Geometry concludes that the series comes from a stochastic system. In this stationary, nonlinear and stochastic series it was more appropriate the model GARCH than the ARMA, because of the smallest error and more adjustment of the prediction curve.

**Discusión.** The results obtained in this work demonstrate the necessity to classify the time series of time before being analyzed, to propose the most appropriate model.

**Conclusión.** The series studied was nonlinear, non stationary, stochastic and in this case the model GARCH fitted better than ARMA.

## **Key Words**

Epidemic prediction, epidemic time series, characterization time series, model GARCH, Epidemiology complex systems

## Introducción

Parece una quimera el momento en que las enfermedades infecciosas pudieran ser eliminadas, aún hasta en los países desarrollados que disponen de mayores recursos para ello. La alta morbilidad asociada con estas enfermedades requiere un alto conocimiento de las especificidades y propagación de las mismas. En Cuba existe un excelente sistema de vigilancia, pero tiene que enfrentar el reto de constantes cambios demográficos y movimientos de poblaciones dentro y fuera del país, entre otros. Existe un eficiente sistema de alerta y respuesta contra la presentación de enfermedades, pero estos programas deben tener en cuenta el presupuesto para las actividades de la salud pública y la complejidad de la propagación; el uso de modelos matemáticos podría apoyar este esfuerzo. (1)

El reto principal es capturar los hechos esenciales de un sistema complejo biológico, tal como lo es una epidemia, en una formulación matemática rigurosa que permita a los investigadores analizar la dinámica espacio temporal, para comprender la aparición, la dispersión de epidemias y las razones de su distribución geográfica, pero lo más importante es apoyar la toma de decisiones con sólidos argumentos científicos.

Es frecuente el análisis y predicción de series de tiempo de datos de enfermos con modelos lineales, tales como “Autoregressive integrated moving average”, (ARIMA), (2) que se manifiesta como: (3)

$$\phi_p(B)\nabla^d \tilde{Z}_t = \theta_q(B)a_t,$$

Donde

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad \theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \quad B\tilde{Z}_t = \tilde{Z}_{t-1}, \quad \tilde{Z}_t = Z_t - \mu \quad \text{y}$$

$$\nabla = 1 - B.$$

Por supuesto, que si un sistema es lineal, entonces los análisis lineales son adecuados, pero si un sistema es no lineal (como suelen ser los sistemas biológicos) entonces deben utilizarse modelos no lineales, la propuesta de este trabajo es que el primer paso para analizar una serie de tiempo de datos epidemiológicos sea que ésta se clasifique previamente, para de esta manera evaluarla con el modelo que mejor le corresponda. Para caracterizar una serie de tiempo es necesario determinar en primer lugar si es estacionaria, luego si es lineal o no y por último si es determinista o estocástica. El objetivo de este trabajo es caracterizar una serie de tiempo de

enfermedades respiratorias agudas en la provincia de Camagüey y evaluarla con un método adecuado a su comportamiento.

## Métodos

Se utilizaron los datos de la serie de tiempo semanal de enfermedad respiratoria aguda reportada entre los años 2000 al 2007 de la provincia de Camagüey, Cuba. En estos años no cambió el sistema de diagnóstico y notificación y en general en el sistema de vigilancia. Se utilizaron las tasas y no números absolutos. Para caracterizar estas series se utilizó el programa Matlab V-7,4-2006. La estacionaridad se evaluó con el método de auto correlación simple (FSA), el comportamiento de linealidad se realizó con la prueba BDS, con el método basado en la geometría espectral se determinó si la serie era estocástica o determinista. . Se comparan los errores en las predicciones con los modelos ARMA y GARCH. y se analizan los errores y las curvas de predicción.

## Resultados.

En la figura 1 se demuestra el comportamiento no estacionario de la serie ya que la función de auto correlación simple (FSA) de la serie original muestra un lento decrecimiento, lo cual es un evidente síntoma de comportamiento no estacionario.. Para llevar esta serie a estacionaria, se utilizó el método de diferenciación; al realizar la primera diferenciación la función de auto correlación simple (figura 2) muestra un abrupto decrecimiento, lo que significa que ya esa serie está transformada en estacionaria. Para confirmar la decisión de que ya la serie está transformada en estacionaria se utilizó el método de Titner's; en este método se calcula la varianza de la original, de la primera y de la segunda diferenciación. En el presente caso la varianza de la segunda diferenciación fue mayor que la de la primera  $33,7 > 12,3$  lo que indica que una diferenciación fue suficiente para convertirla en estacionaria.

Los valores del estadígrafo de la prueba BDS ( $w=8.46$ ) son mayores que los percentiles al 95 y 99 por ciento, por lo que se rechaza la hipótesis nula, es decir se acepta que la serie IRA es no lineal.

La Figura 3 muestra la distribución  $\log(\sigma_i/\text{tr}(\sigma_i))$  (con  $\sigma_i$  los valores propios de la matriz formada con los valores de la serie de orden  $mxn$ ,  $n = 2 \dots 30$  y  $m$  es la parte entera de  $N/n$  con  $N$  la longitud de la serie), de la serie original y la transformada por la geometría espectral, cuando las

dos distribuciones no difieren significativamente se concluye que la serie original proviene de un sistema estocástico, en caso contrario la naturaleza es determinista..

La Figura 4 y 5 muestran los resultados de la predicción a corto plazo utilizando ARMA y GARCH, es evidente que en este caso ( serie no lineal y estocástica), los mejores resultados basados en el error correspondieron al modelo GARCH, el cual fue más adecuado para esta serie...

## Discusión

Es válido el uso de los datos históricos de series de tiempo epidemiológicas porque reflejan la dinámica del comportamiento epidémico con bastante precisión aunque usualmente, se cometan errores de notificación, reportes excesivos o incompletos y otros. (4,5, 6). Los análisis matemáticos y estadísticos de la data histórica han mostrado tener una importancia crucial (7)

Las series de datos epidemiológicas tienen un alto nivel de ruido, son complejas y fuertemente no estacionarias, en los estudios actuales se ha demostrado que la predicción epidemiológica tiene más calidad al utilizarse modelos no lineales ya que estos son capaces de captar factores que contribuyen a una predicción más precisa. (8) Hasta el momento los métodos clásicos de análisis de series de tiempo sólo pueden ser usados en series estacionarias, en las cuales las propiedades estadísticas no varían con el tiempo (9), Se ha encontrado comportamiento no estacionario al evaluar series de tiempo en ritmos cardiacos, (10) , en accidentes de tráfico (11) enfermedades respiratorias agudas y en Malaria (12)

. Existen varias pruebas de hipótesis para evaluara la linealidad de una serie, en este caso se utilizó la prueba BDS descrita por estos tres autores y que ha demostrado ser muy precisa en las distintas evaluaciones que se han realizado en este trabajo (13)

Para determinar si el comportamiento de una serie es estocástica o determinista se ha utilizado previamente el método basado en la geometría espectral el cual se aplicó exitosamente en este estudio ( 14,15 )

Los modelos tradicionales ARIMA se usan con bastante frecuencia (16, 17) pero el análisis y predicción de la dispersión de enfermedades emergentes es difícil de realizar con uno de estos modelos tradicionales (18) los modelos actuales integran la movilidad, locaciones, grupos sociales , cambios demográficos, etc., lo cual aumenta el poder de predicción (19, 20). Se está utilizando con excelentes resultados el modelo para series no lineales GARCH, en series de tiempo de EEG (21), para el control de la fertilidad, y el control de la vacunación oral de la rabia

(22,23,) En los procesos lineales ARIMA tanto la varianza marginal  $Var(z_t)$  como la varianza condicional  $Var(z_t | z_{t-1}, \dots)$  son constantes. Los modelos GARCH son procesos estacionarios que son no lineales en la varianza, aunque su varianza marginal es constante, la varianza condicionada a sus valores pasados no lo es. Como la varianza condicionada representa la incertidumbre de las predicciones, estos procesos van a reflejar un riesgo variable, entendiendo el riesgo como la incertidumbre de las predicciones. Estos procesos fueron propuestos para explicar ciertas propiedades que no pueden explicarse por los modelos ARIMA..

### **Conclusiones**

La serie IRA estudiada resultó ser (a) no lineal, (b) no estacionaria y (c) estocástica y la predicción fue superior en este caso, con el modelo GARCH comparado con el ARIMA.

### **Recomendaciones.**

Se debe ajustar el modelo matemático a la característica de la serie.

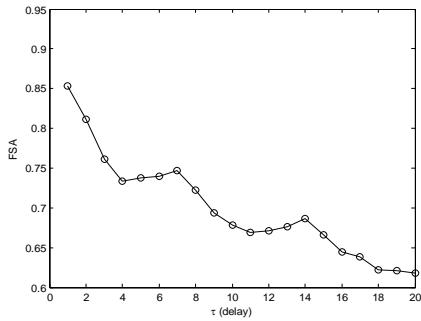
## REFERENCIAS BIBLIOGRÁFICAS

1. Betancourt J A , Ortiz HE, González MA, y Brito HS. Enfoque de los sistemas complejos en la Epidemiología. AMC 2008; Aprobado
2. Box GEP, Jenking GM, Reisel GC. Time series analysis: forecasting and control.. New Jersey: Third edition Prentice Hall 598p; 1991.
3. Peña, D. Estadística. Modelos y métodos. Modelos lineales y series temporales. Madrid: Alianza Editorial S.A. 1993.
4. Trottier H, Philippe P, Roy R. Stochastic modeling of empirical time series of childhood infectious diseases data before and after mass vaccination. *Emerg Themes Epidemiol.* 2006; 3: 9. doi: 10.1186/1742-7622-3-9.
5. Brewster D. Surveillance of measles in Scotland 1983-1991; a comparison of statutory notification and laboratory reporting. *Journal of infection*, 1993; 27(2): 193-201.
6. Torvaldsen S. McIntyre P. Do pertussis notifications reflect incidence or surveillance practices ? A comparison of infant notification rates and hospitalization data in NSW. *New South Wales public health bulletin*. 2003; 14(4): 81-84.
7. Cazelles B, Chavez M, Magny GC, Guégan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. *J R Soc Interface*.2007 ; 4(15):625-636.
8. Cazelles B, Chavez M, Magny GC, Guégan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. *J R Soc Interface*.2007 ; 4(15):625-636.
9. Astolfi L, Cincotti F, Mattia D, Mattiocco M, De Vico Fallani F, Colosimo A, Marciani MG, Hesse W, Zemanova L, Lopez GZ, Kurths J, Zhou C, Babiloni F Estimation of the time-varying cortical connectivity patterns by the adaptive multivariate estimators in high resolution EEG studies. *Conf Proc IEEE Eng Med Biol Soc.* 2006;1:2446-2449.
10. Esposti F, Signorini MG, Ferrario M, Magenes. G.Self-similarity behavior characterization of fetal heart rate signal in healthy and Intrauterine grow retarded fetuses. *Conf Proc IEEE Eng Med Biol Soc.* 2006;1:6157-60
11. Wen J, Yuan P, Deng ZH, Liu KL, Zhang YK, Liu LK, Kong B, Huang SX.Time-series analysis on road traffic injury in China. *Sichuan Da Xue Xue Bao Yi Xue Ban.* 2005; 36 (6):866-869.
12. Stephen Ellner A. Ronald Gallant, and James Theiler. Detecting nonlinearity and chaos in epidemic data Thus nonlinear modeling, and nonlinear forecasting, should be an improvement over linear prediction methods available in URL <http://www.econ.duke.edu/webfiles/arg/papers/newt0.ps> *PLoS ONE.* 2007;2(11):e1181
13. 7H. Andersson, T. Britton, Stochastic Epidemic Models and Their Statistical Analysis. Lecture Notes in Statistics. New York. Springer 2000.

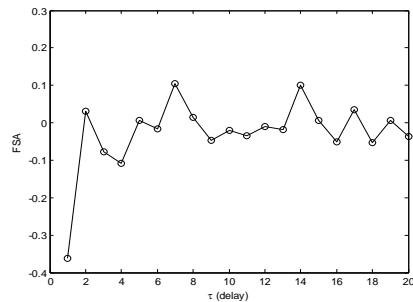
14. Brock, W. Dechert, J. Scheinkma. "A test for independence based on the correlation dimension", *Econometric Rev.* 1996; 15:197-235.
15. J. Bhattacharya, P.P. Kanjilal. "On the detection of determinism in a time series", *Physica D* 132(1999)100–110.
16. HongboXie ,ZhizhongWang, Hai Huang. "Identification determinism in time series based on symplectic geometry spectra", *Physics Letters A* 342 (2005) 156–161.
17. D. Clancy, N. Green. Optimal intervention for an epidemic model under parameter uncertainty *Mathematical Biosciences* 2007; 205: 297–314
18. Rodríguez-Rajo FJ, Valencia-Barrera RM, Vega-Maray AM, Suárez FJ, Fernández-González D, Jato VAnn Prediction of airborne Alnus pollen concentration by using ARIMA models. *Agric Environ Med.* 2006;13(1):25-32
19. Briet OJ, Vounatsou P, Gunawardena DM, Galappaththy GN, Amerasinghe PH. Models for short term malaria prediction in Sri Lanka. *Malar J.* 2008 May 6;7(1):76
20. Duncan JW, Muhamad R, Medina DC., and Dodds PS. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *PNA* August 9, 2005 ;102(32):11157–11162
21. Colizza V, Barrat A, Barthélémy M and Vespignani A.. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study *BMC Medicine* 2007, 5:34 doi:10.1186/1741-7015-5-34
22. Wong KF, Galka A, Yamashita O, Ozaki T.. Modelling non-stationary variance in EEG time series by state space GARCH model. *Comput Biol Med.* 2006 Dec;36 (12):1327-35.
23. Ray T. Sternera,\* , Graham C. Smithb Modelling wildlife rabies: Transmission, economics, and conservation ; *Biological Conservation* 2006; 131163 –179
24. Hyndman, R. J., & Booth, H. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting* 2008;doi:10.1016/j.ijforecast.2008.02.009

## ANEXOS

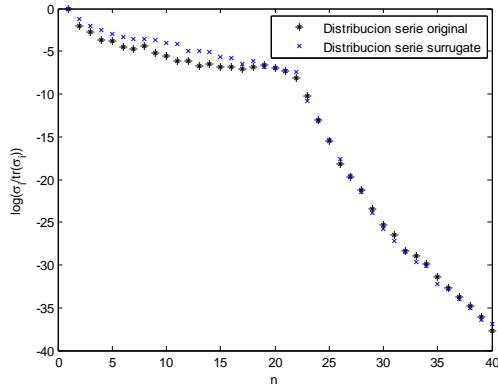
**Fig. 1 Función of auto correlación de la serie original , la cual muestra un lento decrecimiento.**



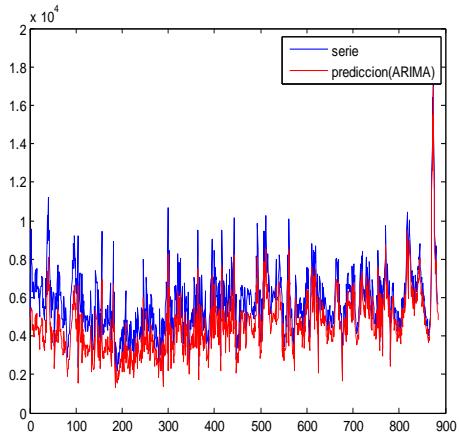
**Fig. 2 Caída abrupta luego de la primera diferenciación.**



**Fig. 3 Prueba basada en la geometría espectral. Este comportamiento demuestra que la serie es estocástica al no diferir significativamente el comportamiento de la distribución de la serie original con respecto a la distribución de la serie transformada.**



**Fig. 4 Predicción al usar ARIMA (Error= 0.7370 )**



**Fig. 5. Predicción al usar GARCH (Error 0.2039)**

