

**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**

**Dr. José A. Betancourt Betancourt<sup>I</sup>; Lic. Humberto Santana Brito<sup>I</sup>; Lic. Eloy Ortiz Hernández<sup>I</sup>; Lic. Nuria Rodríguez Socarrás<sup>II</sup>**

I. Centro de Medicina y Complejidad del Instituto Superior de Ciencias Médicas Carlos J. Finlay. Camagüey, Cuba.

II. Centro Provincial de Higiene y Epidemiología. Camagüey, Cuba.

**RESUMEN**

**Fundamento:** 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.

**Objetivo:** 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 estacionaridad se evaluó con el método de autocorrelación simple, el comportamiento de linealidad se realizó con la prueba diseñada por Brock, Dechert y Scheinkma, 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 autocorrelación simple. 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. Estos resultados demuestran la necesidad

de clasificar las series de tiempo antes de ser analizadas, para proponer el modelo más adecuado.

**Conclusiones:** la serie estudiada resultó ser no lineal, no estacionaria y estocástica y la predicción fue superior en este caso, con el modelo GARCH comparado con el ARMA.

**DeCS:** Predicción epidemiológica, series de tiempo epidemiológica, caracterización series de tiempo, modelo GARCH, epidemiología sistemas complejos.

## ABSTRACT

**Background:** in Cuba an excellent system of surveillance exists, but has to face the challenge of constant demographic changes and populations' movements inside and outside of the country, among other. An efficient system of alert and answer against the presentation of diseases exists, but these programs may take into account the budget for the public health activities and the complexity of the propagation; the use of mathematical models could support this effort.

**Objective:** to characterize a series of time of acute respiratory diseases in Camagüey province and to evaluate it with an appropriate method to its behavior.

**Methods:** rates of series of time weekly of acute respiratory diseases were used, reported between the years 2000 to 2007 of Camagüey province, Cuba. To characterize these series Matlab V-7,4-2006 program was used. The stationarity was evaluated with the simple autocorrelation method, the linearity behavior was carried out with the test designed by Brock, Dechert and Scheinkma, with the method based on the spectral geometry that was determined if the series was stochastic or determinist. Errors were compared in the predictions with ARMA and GARCH models, errors and prediction curves were analyzed.

**Results:** the non stationary behavior of the series is demonstrated, based on the simple autocorrelation function. The test indicates that the studied series was not lineal. The spectral Geometry test concluded that the series comes from a stochastic system. In this stationary, not lineal and stochastic series was more adequate GARCH model than ARMA, given the smallest error and more adjustment of the prediction curve. These results demonstrated the necessity to classify the series of time before being analyzed, to propose the most appropriate model.

**Conclusions:** the studied series turned out to be not lineal, no stationary and stochastic and the prediction was superior in this case, with the model GARCH compared with ARMA.

**DeCS:** 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. <sup>1</sup>

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), <sup>2</sup> que se manifiesta como: <sup>3</sup>

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.

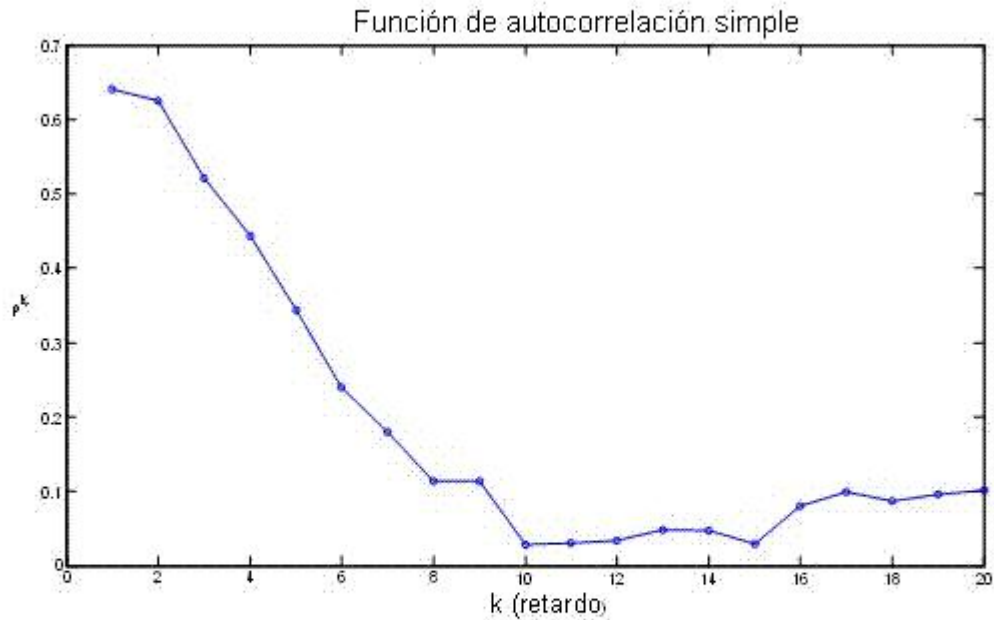
El objetivo de este trabajo es caracterizar una serie de tiempo de enfermedades respiratorias agudas en la provincia de Camagüey y evaluarla con un método adecuado a su comportamiento.

## MÉTODO

Se utilizaron los datos de la serie de tiempo semanal de enfermedad respiratoria aguda reportada entre los años 2000 y 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 autocorrelación simple (FSA), el comportamiento de linealidad se realizó con la prueba diseñada por Brock, Dechert y Scheinkma (BDS), con el método basado en la geometría espectral se determinó si la serie era estocástica o determinista. Se compararon los errores en las predicciones con los modelos ARMA y GARCH y se analizaron los errores y las curvas de predicción.

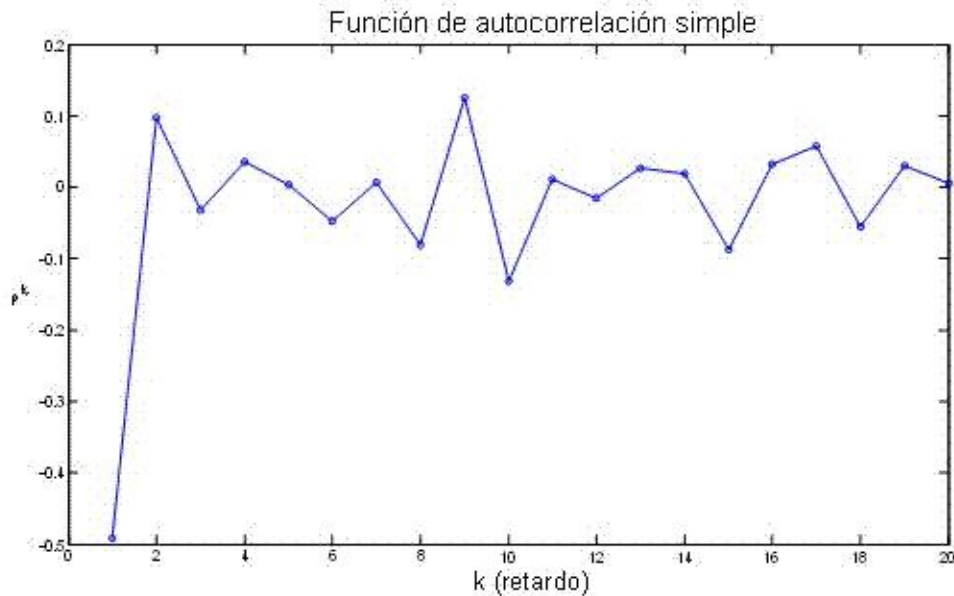
## RESULTADOS

Con respecto al comportamiento no estacionario de la serie la función de autocorrelación simple (FSA) de la serie original mostró un lento decrecimiento, lo cual es un evidente síntoma de comportamiento no estacionario. Figura1



**Fig. 1. Función de autocorrelación de la serie original que muestra un lento decrecimiento.**

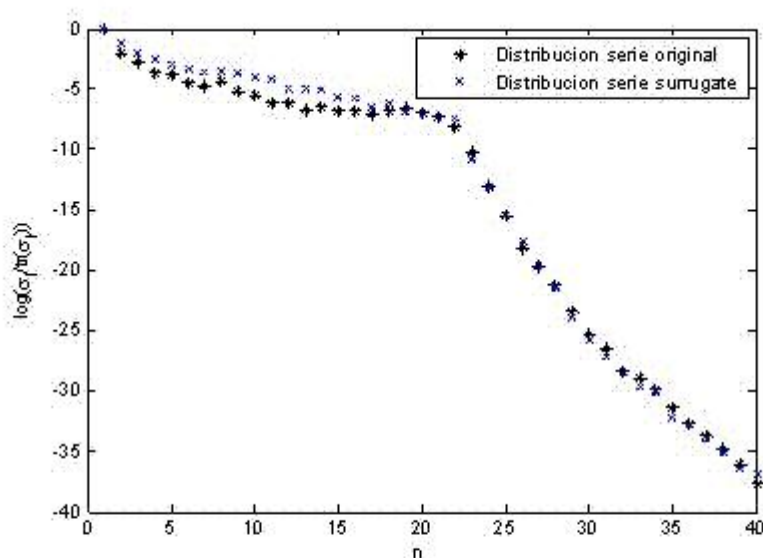
Para llevar esta serie a estacionaria, se utilizó el método de diferenciación; al realizar la primera diferenciación la función de autocorrelación simple mostró un abrupto decrecimiento, lo que significa que ya esa serie está transformada en estacionaria. Para confirmar esta decisión se utilizó el método 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 indicó que una diferenciación fue suficiente para convertirla en estacionaria. Figura 2



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

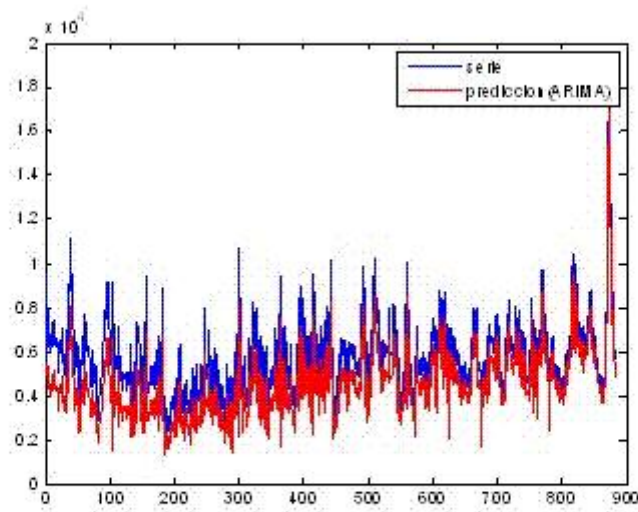
Los valores del estadígrafo de la prueba BDS ( $w=8,46$ ) son mayores que los percentiles al 95 y 99 %, por lo que se rechaza la hipótesis nula, es decir se acepta que la serie IRA es no lineal.

La distribución  $\log(s_i/tr(s_i))$  (con  $s_i$  los valores propios de la matriz formada con los valores de la serie de orden  $m \times n$ ,  $n = 2...30$  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. Figura 3

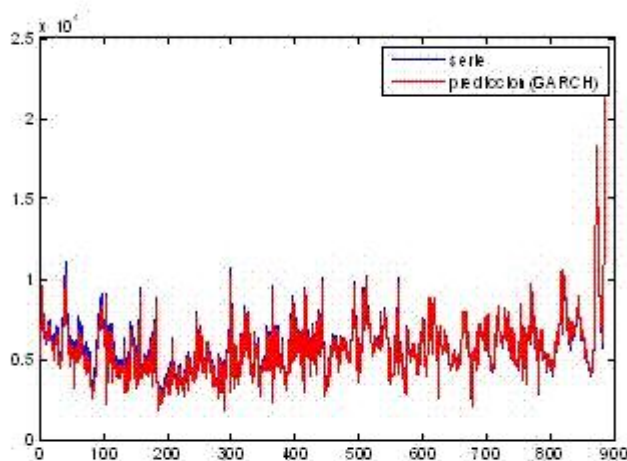


**Fig. 3. Prueba basada en la geometría espectral que demuestra que la serie es estocástica al no diferir significativamente el comportamiento de la distribución de la serie original con respecto a la distribución de la serie transformada.**

Con relación a los resultados de la predicción a corto plazo utilizando ARMA y GARCH, resultó 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. Figuras 4 y 5



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



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

## 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.<sup>4-6</sup> Los análisis matemáticos y estadísticos de la data histórica han mostrado tener una importancia crucial.<sup>7</sup>

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.<sup>8</sup> 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<sup>9</sup>. Se ha encontrado comportamiento no estacionario al evaluar series de tiempo en ritmos cardiacos,<sup>10</sup> en accidentes de tráfico<sup>11</sup> enfermedades respiratorias agudas y en malaria.<sup>12</sup>

Existen varias pruebas de hipótesis para evaluar 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.<sup>13</sup>

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.<sup>14, 15</sup>

Los modelos tradicionales ARIMA se usan con bastante frecuencia,<sup>16, 17</sup> pero el análisis y predicción de la dispersión de enfermedades emergentes es difícil de realizar con uno de estos modelos tradicionales,<sup>18</sup> los modelos actuales integran la movilidad, locaciones, grupos sociales, cambios demográficos, etc., lo cual aumenta el poder de predicción.<sup>19, 20</sup> Se está utilizando con excelentes resultados el modelo para series no lineales GARCH, en series de tiempo de EEG<sup>21</sup>, para el control de la fertilidad, y el control de la vacunación oral de la rabia.<sup>22, 23</sup> En los procesos lineales ARIMA tanto la varianza marginal  $Var(z_1)$  como la varianza condicional  $Var(z_1 / z_{t-1}, \dots)$  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.

La serie insuficiencia respiratoria aguda estudiada resultó ser no lineal, no estacionaria y 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 las características de la serie.

## REFERENCIAS BIBLIOGRÁFICAS

1. Betancourt JA, Ortiz HE, González MA, Brito HS. Enfoque de los sistemas complejos en la Epidemiología. Disponible en: <http://www.sld.cu/sitios/complejidad>

2. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. New Jersey: Third edition Prentice Hall; 1991.p. 598.
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.
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. *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):625-36.
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):625-36.
9. Astolfi L, Cincotti F, Mattia D, Mattiocco M, DeVico Fallani F, Colosimo A. Estimation of the time-varying cortical connectivity patterns by the adaptive multivariate estimators in high resolution EEG studies. *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 retardated fetuses. *Eng Med Biol Soc* 2006; 1:6157-60.
11. Wen J, Yuan P, Deng ZH, Liu KL, Zhang YK, Liu LK. Time-series analysis on road traffic injury in China. *Sichuan Da Xue Xue* 2005; 36(6):866-869.
12. Stephen Ellner A, Gallant R, Theiler J. Detecting nonlinearity and chaos in epidemic data Thus nonlinear modeling, and nonlinear forecasting, should be an improvement over linear prediction methods. *PLoS ONE* 2007; 2(11):1181.
13. Andersson H, Britton T. 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. *Rev Econometric* 1996; 15: 197-235.
15. Bhattacharya J, Kanjilal PP. On the detection of determinism in a time series. *Physica J* 1999; 132(2):10.
16. Hongbo Xie, ZhizhongWang Hai Huang. Identification determinism in time series based on symplectic geometry spectra. *Physics Letters A* 2005; 342 (12):156-61.
17. Clancy D, Green N. 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; 7(1):76.



20. Duncan JW, Muhamad R, Medina DC, Dodds PS. Multiscale, resurgent epidemics in a hierarchical metapopulation model. BMC Medicine 2005; 102(32):11157162.
21. Colizza V, Barrat A, Barthélemy M, Vespignani A. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. BMC Medicine 2007; 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; 36(12):1327-35.
23. Sternera Ray T, Graham C. Smithb modelling wildlife rabies: transmission, economics, and conservation. Biological conservation 2006; 13:1163-179.

Recibido: 16 de septiembre de 2008.

Aceptado: 5 de diciembre de 2008.

*José A. Betancourt Betancourt.* Médico veterinario. Master dedicado a la gestión de proyectos biomédicos y sistemas complejos. Centro de Medicina y Complejidad del Instituto Superior de Ciencias Médicas Carlos J. Finlay. Camagüey, Cuba.