

¿Aleatoriedad o determinismo no lineal? Análisis de series temporales de contaminantes atmosféricos urbanos

J.H. Lucio

Departamento de Física, Universidad de Burgos, Escuela Politécnica Superior,
Avda. Cantabria s/n, 09006 Burgos, España
e-mail: jlucio@ubu.es

C. Caballero

Departamento de Física, Universidad de Burgos, Facultad de Ciencias,
Plaza Misael Bañuelos s/n, 09001 Burgos, España

Recibido el 7 de agosto de 2004; aceptado el 3 de enero de 2005

El análisis clásico de series temporales se ha basado principalmente en un enfoque estocástico lineal (se suponía que la principal estructura de una serie era la autocorrelación lineal en diferentes retardos). En los últimos años han cobrado importancia los sistemas caóticos: una señal compleja puede explicarse en ocasiones por unas ecuaciones deterministas, generalmente no lineales. Este tipo de sistemas aparece en muchos ámbitos de la física y la ingeniería, también en la contaminación atmosférica. En este artículo analizamos series de los tres principales gases implicados en la contaminación fotoquímica: NO, NO₂ y O₃. Nuestro objetivo es aclarar si la estructura principal de estas series es estocástica y lineal o determinista y no lineal. Para ello tratamos de construir un espacio de inserción equivalente al espacio de fases del supuesto sistema determinista del que proceden los datos. En las tres series detectamos que existe una estructura determinista de baja dimensión, pero es más fuerte el componente estocástico.

Descriptores: Series temporales no lineales; caos; series temporales; ozono troposférico.

Classical analysis of time series has been founded mainly on a linear stochastic approach (the main structure of a series was supposed to be the linear autocorrelation in several lags). In latter years chaotic systems have acquired importance: a complex signal can sometimes be explained by some deterministic, usually nonlinear, equations. This kind of systems is found in many areas of Physics and Engineering, also in atmospheric pollution. In this paper we analyze series of the three main gases involved in photochemical pollution: NO, NO₂ and O₃. Our aim is to clarify if the main structure of these series is stochastic and linear or deterministic and nonlinear. For this purpose we try to build an embedding space equivalent to the phase space of the hypothetic deterministic system from which the series are measured. In the three series we have found that a low-dimensional deterministic structure exists, but the stochastic component is stronger.

Keywords: Nonlinear time series; chaos; time series; tropospheric ozone.

PACS: 05.45.Tp; 47.52+j; 95.10.Fh; 82.50.Nd

1. Introducción

Un serio problema ambiental en muchas ciudades industrializadas es la contaminación fotoquímica, en la que óxidos de nitrógeno reaccionan con hidrocarburos volátiles en presencia de luz solar, dando lugar a diversos contaminantes secundarios, de los que el más representativo es el ozono. Estos contaminantes secundarios, conocidos genéricamente como *oxidantes fotoquímicos*, son los causantes cada año de cientos de muertes y muchos más casos de problemas respiratorios severos.

Para explicar la estructura de las series temporales de los contaminantes atmosféricos podemos plantear dos opciones extremas: suponer que responden a un modelo estocástico lineal o que son manifestación de algún sistema determinista no lineal. Trataremos de responder a la pregunta ¿cuál es la estructura predominante en cada serie?

2. Datos experimentales

Disponemos de valores de las concentraciones de NO, NO₂ (en adelante NO₂) y O₃ (O₃), expresadas en $\mu\text{g}/\text{m}^3$, medidas

a nivel del suelo en el centro histórico de la ciudad de Burgos, España. Las concentraciones de NO y NO₂ se midieron con un analizador de quimioluminiscencia y las de O₃ con un fotómetro de absorción de luz ultravioleta. Todos los datos

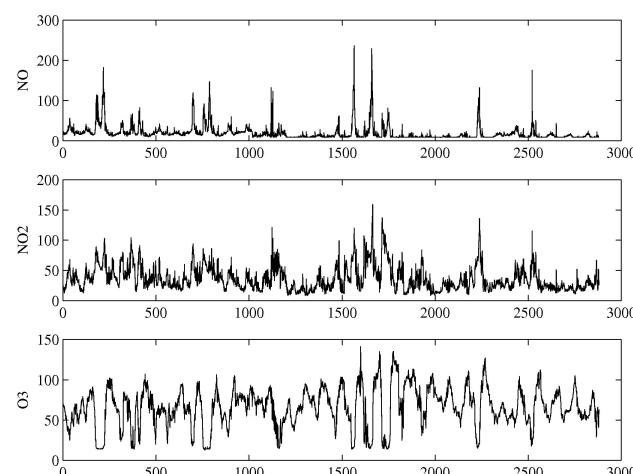


FIGURA 1. Las tres series originales.

tienden un periodo de muestreo de 15 minutos y corresponden al mismo intervalo de tiempo: el mes de junio de 1998, por lo que cada serie contiene 2880 valores.

En la Fig. 1 se representan las tres series frente al tiempo, se puede apreciar una cierta estacionalidad de periodo 96 (1 día), especialmente en O3.

3. Sistemas caóticos. Espacio de inserción

Un sistema determinista puro evoluciona en el tiempo de una forma posiblemente compleja, pero que viene fijada por unas ecuaciones (en casos sencillos son ecuaciones diferenciales lineales). Si el sistema es continuo (flujo) $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$, y si es discreto (mapa) $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$, donde \mathbf{x} , denominado vector de estado, es un vector de dimensión m . El espacio de las \mathbf{x} recibe el nombre de espacio de fases o espacio de estados. El estudio de la evolución de los sistemas en el espacio de fases ha llevado a importantes avances en el conocimiento del caos.

Una función de medida, $s(\cdot)$, puede dar una variable (pongamos escalar y real) a partir de cada valor del vector de estado \mathbf{x} . Si el sistema es un flujo y la función de medida se aplica a intervalos regulares de tiempo hablamos de *muestreo*, y la sucesión obtenida recibe el nombre de *serie temporal*, $\{s_n = s(\mathbf{x}_n)\}$.

Los teoremas de inserción (*embedding theorems*) permiten inferir las propiedades del sistema a partir de una serie temporal medida de él (véanse, por ejemplo, las citas en las Refs. 1 y 2). En concreto, dichos teoremas garantizan que con una serie de tamaño infinito y en ausencia de ruido es posible reconstruir un espacio vectorial *equivalente* al espacio de estados original. Una técnica muy utilizada es la *reconstrucción de retardos*, que consiste en construir vectores

$$\mathbf{s}_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \dots, s_{n-\tau}, s_n), \quad (1)$$

a partir de la serie temporal. El espacio generado por los vectores \mathbf{s}_n se conoce como *espacio de inserción* (*embedding space*) y los enteros m y τ reciben los nombres de *dimensión de inserción* y *retardo*, respectivamente.

Para estimar el retardo τ se suele emplear el método de la *información mutua*, basado en la entropía de Shannon. Se define como la información que se tiene sobre $s(t + \tau)$ si se conoce $s(t)$. Se puede estimar construyendo un histograma univariante de la serie $\{s(t)\}$ y otro bivariante de $\{s(t), s(t + \tau)\}$. Si p_i es la probabilidad de que $s(t)$ esté en la caja i y $p_{ij}(\tau)$ la probabilidad de que $s(t)$ esté en i y $s(t + \tau)$ esté en j , se puede estimar la información mutua así:

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln p_{ij}(\tau) - 2 \sum_i p_i \ln p_i. \quad (2)$$

Se aconseja tomar como estimación de τ el valor que da el primer mínimo en la información mutua. En la Fig. 2 se muestran las gráficas de la información mutua para las tres series, calculada hasta el retardo 100 y con 50 cajas en los histogramas [1]. Los valores propuestos para τ (señalados con flecha) son: 50 para NO y 30 para NO2 y O3.

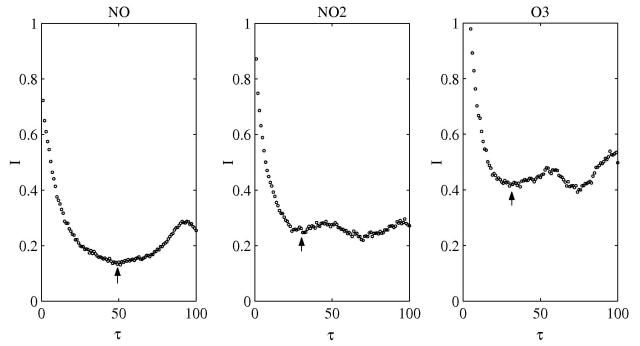


FIGURA 2. Información mutua para las tres series en función del retardo τ .

Para estimar la dimensión de inserción, m , existen numerosos procedimientos, siendo el más empleado el método de los *vecinos próximos falsos* (en inglés *false nearest neighbors*) [1]. Se basa en que dos puntos lejanos del espacio de estados pueden parecer próximos en alguna proyección determinada, es decir, si la dimensión del espacio es insuficiente para desdoblarse todas las características geométricas del sistema. Si vamos aumentando la dimensión de la proyección irán desapareciendo esos falsos vecinos. Se halla el vecino más próximo a cada punto s_i , llamémoslo s_j , y se calcula la fracción

$$R_i = \frac{|s_{i+1} - s_{j+1}|}{\|s_i - s_j\|}, \quad (3)$$

que marca el alejamiento del vecino más próximo al aumentar la dimensión en la unidad. Si la fracción es superior a un umbral prefijado rf (del inglés *ratio factor*), entonces se toma s_j como vecino falso. Finalmente, se calcula la fracción de los vecinos más próximos falsos (en adelante FFNN, de *fraction of false nearest neighbors*) en función de la dimensión. Cuando esa fracción se hace nula a partir de un valor de m se puede pensar que se ha encontrado la dimensión que desdobra completamente el atractor.

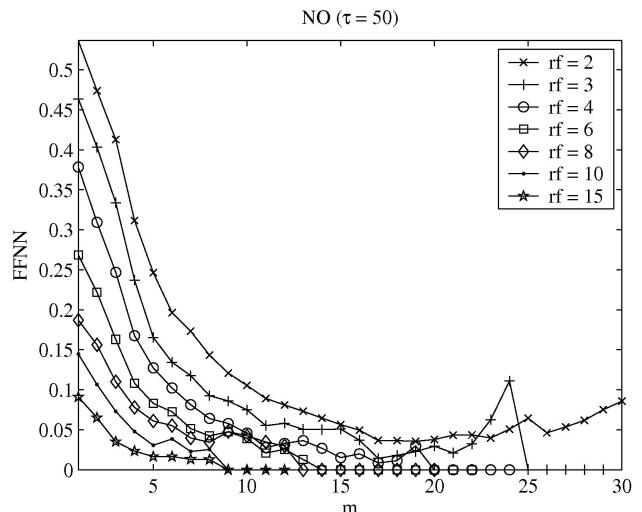
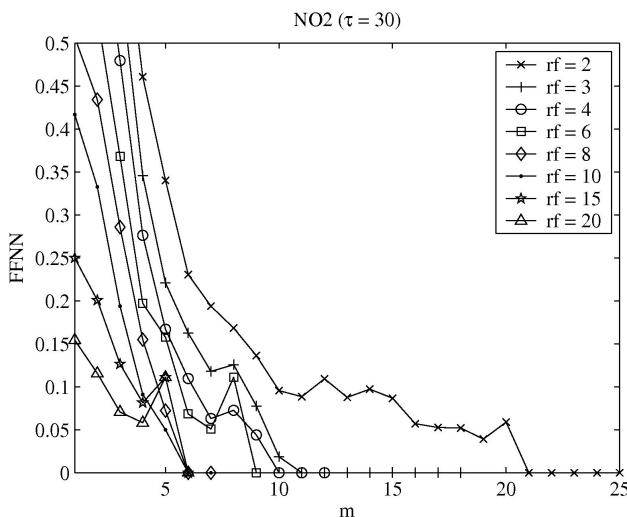
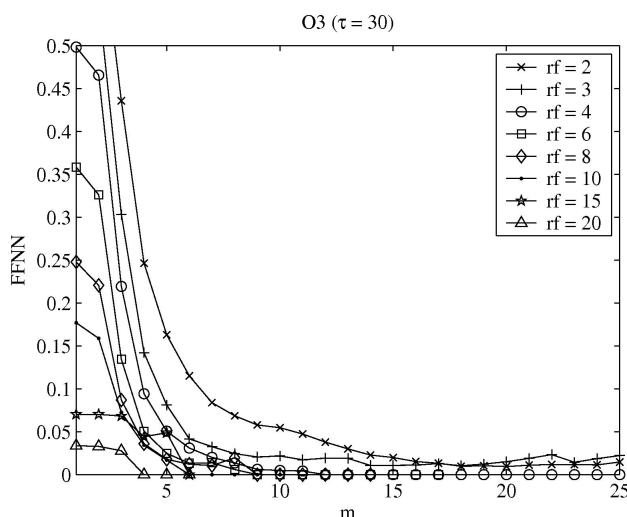
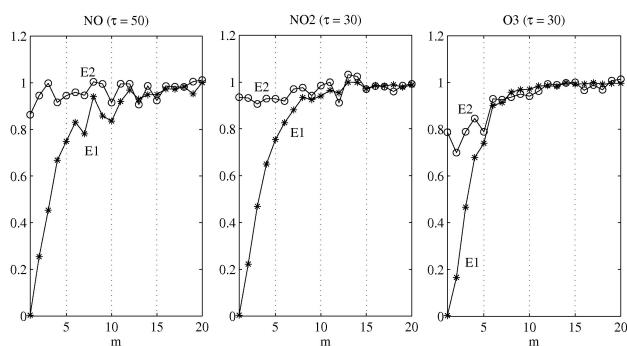


FIGURA 3. FFNN en función de m , para NO.

FIGURA 4. FFNN en función de m , para NO2.FIGURA 5. FFNN en función de m , para O3.FIGURA 6. $E1(m)$ y $E2(m)$ para NO, NO2 y O3.

Calculamos la FFNN en función de la dimensión de inserción m , para diferentes valores del umbral rf (Figs. 3, 4 y 5) [3].

Los menores valores de rf parecen relacionados con el ruido (señal estocástica o determinista de dimensión eleva-

da). A medida que aumenta rf se reduce la dimensión de inserción, siendo 9, 6 y 4 los valores propuestos para NO, NO2 y O3, respectivamente.

Existen otros muchos métodos para la estimación de la dimensión de inserción. El método de Cao [4] incluye dos funciones de m , $E1(m)$ y $E2(m)$. $E1(m)$, similar a FFNN, sirve para estimar la dimensión de inserción y tiene la ventaja de no necesitar umbrales, como los que necesita FFNN para el cociente (3). En particular, si la señal es determinista, $E1(m)$ crece rápidamente para las primeras dimensiones y se estabiliza bruscamente en un valor de m , que es la estimación de la dimensión de inserción. Si la señal es estocástica $E1(m)$ crece monótonamente, no presenta una saturación brusca. $E2(m)$ permite distinguir una señal determinista de una estocástica, ya que mide la dilatación (al aumentar la dimensión en la unidad) de la distancia media entre los valores de la serie correspondientes al próximo elemento de los vectores de retardo y sus vecinos más próximos. En particular, si la señal es determinista (y posee un atractor), $E2(m)$ se asemeja mucho a $E1(m)$: comienza muy baja y sube rápidamente hasta que se estanca a partir de la dimensión de inserción. Para una señal estocástica, $E2(m)$ permanece invariable desde el comienzo (entorno a 1), independientemente de m .

En la Fig. 6 se representan $E1(m)$ y $E2(m)$ para las tres series. $E1(m)$ podría reflejar un débil comportamiento determinista, al menos en O3, mientras que $E2(m)$ permanece prácticamente invariable desde el principio entorno a la unidad, lo que parece indicar que las tres series (menos claramente en O3) son predominantemente estocásticas.

Por último, el test no linealidad de Casdagli [5] (no mostrado) señala las tres series como predominantemente lineales (con el patrón típico de una serie estocástica lineal).

4. Conclusiones

Hemos comprobado que las series de NO, NO2 y O3 (al menos en el intervalo temporal del orden de un mes y con periodo de muestreo cercano a 15 minutos) presentan simultáneamente un carácter determinista, con dimensión de inserción no muy elevada, y un comportamiento estocástico, al parecer lineal. De los dos componentes parece predominar el estocástico y el carácter no lineal parece ser débil.

Es bien sabido que este tipo de series temporales admite generalmente la representación con modelos estocásticos lineales (como los ARMA [6]) pero, a la vista de los resultados, parecen también ofrecer —al menos el O3— una posibilidad de análisis determinista no lineal. La complejidad (en el sentido de que tienen mezcladas estructuras muy diversas) de las series queda patente en los análisis practicados.

Agradecimientos

Los autores desean agradecer a la Junta de Castilla y León su gentileza en la donación de los datos y al árbitro su paciencia y profesionalidad al corregir errores en el trabajo.

1. H.D.I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer 1996).
2. H. Kantz y T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press 1997).
3. R. Hegger, H. Kantz y T. Schreiber, *CHAOS* **9** (1999) 413.
4. L. Cao, *Physica D* **110** (1997) 43.
5. M. Casdagli, *J. Roy. Stat. Soc.* **54** (1991) 303.
6. G.E. Box, G.M. Jenkins y G.C. Reinsel, *Time Series Analysis. Forecasting and Control*, 3rd edition (Prentice Hall 1994).