

## **Influencia del Tratamiento de Datos en la Detección de Regímenes Caóticos en series Temporales**

**Matías Rafti , Cristina Cordero y José L. Vicente**

Universidad Nacional de La Plata, Facultad de Ciencias Exactas, Dpto. de Química, INIFTA, CICPBA, CONICET, Casilla de Correo 16, Sucursal 4, (B1904DPI), La Plata-Argentina  
(e-mail: mrafti@quimica.unlp.edu.ar)

*Recibido Sept. 15, 2009; Aceptado Dic. 02, 2009; Versión Final recibida Dic. 17, 2009*

---

### **Resumen**

Se estudia la influencia del procedimiento de filtrado aplicado sobre series temporales para su uso en la caracterización y detección de regímenes no-lineales. Para esto, se toma como ejemplo de dichos métodos, un algoritmo ampliamente utilizado para el cálculo de la dimensión de correlación de la trayectoria en el espacio de las fases (el algoritmo de Grassberger-Procaccia). El interés de este ejemplo de estudio radica en la similitud con el procedimiento que se aplica al analizar imágenes experimentales de sistemas fisicoquímicos de reacción-difusión, provenientes de técnicas de análisis superficial como la microscopía de emisión de fotoelectrones. Los resultados de las simulaciones realizadas muestran como el uso de un parámetro de filtrado inadecuado puede conducir a caracterizar erróneamente estados como no lineales o caóticos.

*Palabras clave: series temporales, tratamiento de datos, caos, dimensión de correlación, fractales*

## **Influence of Data Handling on the Detection of Chaotic Regimes in time series Analysis**

### **Abstract**

The influence of the filtering scheme on time series for their use in the characterization and detection of non linear regimes. As an example, the Grassberger-Procaccia algorithm for phase-space trajectory correlation dimension is used. The main interest of such analysis is to study the similarities with the processing of experimental images from reaction-diffusion systems obtained via surface science standard tools such as photoelectron emission microscopy. Simulation results show how inadequate filtering parameter choice may lead to erroneous characterization of systems as non-linear or chaotic.

*Keywords: time series, data treatment, chaos, correlation dimension*

## INTRODUCCIÓN

Existe una variedad de sistemas fisicoquímicos donde pueden observarse comportamientos complejos no lineales, como por ejemplo la oscilación temporal de velocidades de reacción y cubrimientos superficiales en sistemas de reacción-difusión (Imbihl y Ertl 1995, Rafti et al., 2007). Las oscilaciones observadas pueden ser de distinta naturaleza (regulares o irregulares) y su análisis exhaustivo permite revelar aspectos generales de la dinámica del sistema que las produce. El comportamiento irregular observado en algunos casos puede deberse a la existencia de procesos subyacentes de naturaleza estocástica (innumerables grados de libertad) o de origen puramente determinista (número finito de grados de libertad); esto es, gobernados por ecuaciones diferenciales no lineales altamente sensibles a las condiciones iniciales. Los sistemas que responden a esta última descripción se denominan caóticos (Makarenko 2009). El problema que surge es cómo distinguir a partir de la evolución temporal de un sistema, caos determinístico de ruido puramente estocástico debido a que la señal medida siempre tendrá un cierto nivel de ruido aleatorio (Argyris y Andreadis 1997 y Rosso et al., 2007). La teoría del caos provee algunas herramientas para llevar a cabo este análisis, entre las cuales se encuentra el cálculo de la dimensión de correlación del atractor del sistema, que a su vez está directamente relacionado con el número de grados de libertad activos (Osborne y Provenzale 1989, Ramdani et al., 2007).

Al estudiar mediante simulaciones numéricas sistemas extendidos en el espacio y en el tiempo, que pueden presentar dinámicas caóticas (como los sistemas de reacción-difusión), se aplican modelos que utilizan algún tipo de discretización. Las mismas consisten en dividir el sistema macroscópico en subunidades o celdas mesoscópicas, que retienen las características dinámicas (donde es válida la hipótesis de campo medio), para luego acoplarlas con reglas que reflejen adecuadamente los mecanismos de difusión entre las mismas (Zhdanov 2004, Rafti et al., 2006, Hongler et al., 2009). Ahora bien, resulta esencial la elección del tamaño de las celdas dado que, un error por exceso o por defecto tiende, o bien a suavizar la dinámica no permitiendo observar fenómenos en la escala de interés o a indicar acoplamientos y correlaciones inexistentes. La situación se torna crítica cuando se comparan simulaciones con resultados experimentales, por ejemplo en el caso de imágenes obtenidas vía Microscopía de Emisión de Fotoelectrones (PEEM), además de tener un valor límite de resolución fijado por la técnica, son modificadas por la utilización de algún filtros (Uecker 2007).

En el presente trabajo se estudian los efectos de la aplicación de un filtro estándar sobre series temporales generadas a partir de un proceso estocástico (ruido blanco); se demuestra como el uso de un filtro inadecuado puede transformar la señal original en un proceso que no cumple con la condición de presentar iguales características a distintas escalas; esto es, una señal inicialmente estocástica puede aparecer por efecto del filtrado y de modo artificial como altamente correlacionada.

## METODOLOGÍA

### *Dimensión de correlación de un atractor y algoritmo de Grassberger-Procaccia*

Una de las técnicas utilizadas para caracterizar sistemas que presentan comportamiento caótico la llamada reconstrucción del atractor en el espacio de las fases (Kantz y Schreiber 2003). Dado un sistema dinámico cuyo estado se define por medio de un vector,

$$X \in R^n, \quad (1)$$

y cuya evolución con el tiempo  $t$  está gobernada por un campo vectorial de acuerdo a,

$$f : R^n \rightarrow R^n, \frac{dX}{dt} = f(X), \quad (2)$$

el número de coordenadas  $n$  en el vector de estado, caracteriza el número máximo posible de grados de libertad del sistema y al espacio  $n$ -dimensional en donde se representa a  $X$  se lo denomina espacio de las fases. Si  $f$  no depende del tiempo, el sistema es autónomo, las soluciones del sistema

se caracterizarán por un estado estacionario que debe ser acotado a medida que  $t \rightarrow \infty$ , por lo tanto la trayectoria debe tender a circunscribirse en un subconjunto de puntos del espacio de las fases (de dimensión menor o igual a  $n$ ), este subconjunto se denomina atractor. Para ciertos tipos de  $f$  no lineales, muy sensibles a cambios en las condiciones iniciales, el sistema se denominará caótico y el atractor resultará un objeto fractal, i.e., de dimensión topológica no entera (Akay 2001). Ahora bien, dada una serie temporal unidimensional, considerada como la única componente conocida de un vector  $n$ -dimensional desconocido, el problema que se plantea es reconstruir una trayectoria que sea topológicamente equivalente al atractor real en el espacio de las fases. Para lograr esto, se usa el esquema de Takens (Packard et al., 1980) de coordenadas retardadas, que consiste en proponer un tiempo de retraso adecuado  $\tau$  y construir vectores a partir de la serie temporal discreta  $s(t_i)$  en un espacio  $M$ -dimensional artificial,

$$\vec{x}_i = [s(t_i), s(t_i + \tau), s(t_i + 2\tau), \dots, s(t_i + (M-1)\tau)]. \quad (3)$$

Una vez establecida la trayectoria en el espacio de las coordenadas retardadas, se puede evaluar la dimensión de correlación del atractor ( $D_2$ ) mediante el algoritmo de Grassberger y Procaccia (GP) (Harikrishnan y col. 2006). Básicamente,  $D_2$  es una extensión del cálculo de la tradicional dimensión fractal  $D_0$  (que simplemente calcula el número de "cajas" de tamaño uniforme necesarias para cubrir el atractor). La dimensión de correlación toma en cuenta también cuan densamente se agrupan los puntos del atractor unos alrededor de los otros; para su cálculo primero debe obtenerse el número relativo de puntos dentro de un radio  $R$  medido desde un punto  $i$ -ésimo elegido aleatoriamente,

$$p_i(R) = \lim_{N_v \rightarrow \infty} \frac{1}{N_v} \sum_{j=1, j \neq i}^{N_v} H(R - \|\vec{x}_i - \vec{x}_j\|), \quad (4)$$

$N_v$  es el número total de vectores reconstruidos,  $H$  es la función escalón de Heaviside y  $\|\dots\|$  representa la distancia euclídea en  $R^M$ . Al promediar esta cantidad sobre  $N_c$  puntos elegidos se obtiene la función de correlación,

$$C_M(R) = \frac{1}{N_c} \sum_i^{N_c} p_i(R). \quad (5)$$

Se define la dimensión de correlación como el índice de escala de la variación de  $C_M(R)$  con  $R$  a medida que  $R \rightarrow \infty$ , esto es,

$$D_2(M) \equiv \lim_{R \rightarrow \infty} \frac{d \log C_M(R)}{d \log R}. \quad (6)$$

Al estudiar la variación de  $D_2$  con  $M$  se observa que en sistemas estocásticos aleatorios (ruido blanco),  $D_2$  no alcanza un valor límite al incrementar  $M$  (i.e., la trayectoria ocupa todo el espacio para cualquier valor de  $M$ ), mientras que para sistemas con dinámicas caóticas se alcanza una saturación de  $D_2$  al alcanzar  $M$  el valor de la dimensión del atractor característico del sistema. Sin embargo, cierto tipo de sistemas estocásticos (ruidos coloreados), deben analizarse con especial cuidado dado que pueden dar como resultado dimensiones de correlación finitas y pueden inducir a conclusiones erróneas (Osborne y Provenzale 1989).

#### *Series temporales y algoritmo de filtrado*

Las series temporales se construyeron utilizando un generador de números aleatorios uniformemente distribuidos del paquete de software MATLAB® (procesador 2.67GHz, 2.96 GB RAM), a las que luego se les aplicó un filtro de convoluciones con ventanas de distinto ancho, es decir se aplicó un algoritmo que consistió en: a) seleccionar un parámetro  $nff$  (asociado al ancho de la ventana), b) seleccionar el primer punto de la serie sin filtrar  $s(t_i)$ ,  $i=nff+1$ , c) sumar y promediar los valores desde  $i=nff-1$  hasta  $i=nff+1$  y normalizar dividiendo por el ancho de la ventana ( $2*nff+1$ ). El resultado de

este procedimiento es una serie  $s_f(t_i)$ , con  $N-2*nff$  puntos a partir de una  $s(t_i)$  de  $N$  puntos. Como es usual, en cada caso se calcula el tiempo de retraso  $\tau$  para  $s_f(t_i)$  evaluando el punto en que la función de autocorrelación decae hasta  $e^{-1}$  (Akay, 2001). En la figura 1 se muestran una serie temporal sin filtrar, una filtrada y un diagrama que ilustra el esquema de filtrado utilizado.

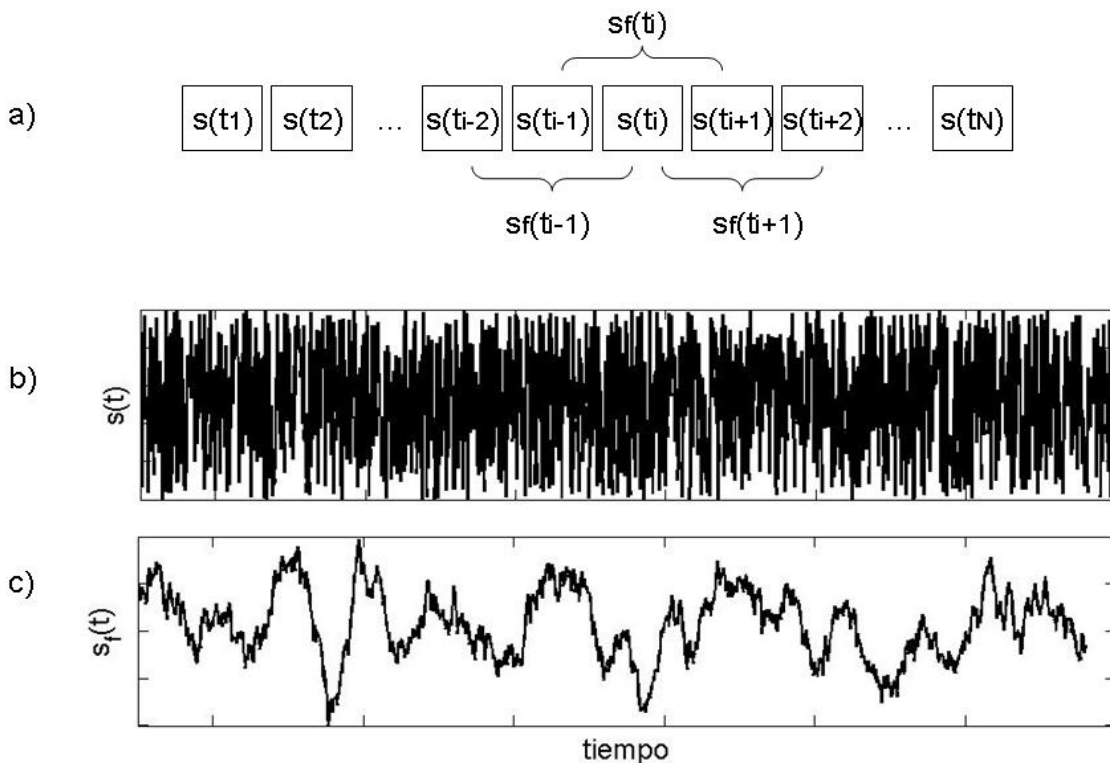


Fig. 1: a) Diagrama del esquema de filtrado utilizado para  $nff=1$ . b) Ejemplo de serie temporal utilizada en el análisis sin filtrar. c) Serie temporal filtrada. Longitud total de la serie  $N=2 \times 10^4$  (se muestra un extracto de  $2 \times 10^3$  unidades temporales) y parámetro de filtrado  $nff=80$ .

## RESULTADOS Y DISCUSIÓN

Para evaluar la influencia del tamaño de las series temporales en el análisis y caracterización del atractor a través de su dimensión de correlación,  $D_2(M)$ , se realizaron pruebas con valores de  $N = 1, 2$  y  $5 \times 10^4$  puntos, para las que se usaron parámetros de filtrado  $nff=40, 80$  y  $200$  respectivamente. Los resultados de estas simulaciones se muestran en la figura 2, donde se puede observar que para valores  $M > 6$  se alcanza un valor de convergencia de  $D_2(M)$ , y que un incremento de  $N$  por encima de  $2 \times 10^4$  no cambia los resultados observados. Se muestra también el comportamiento de  $D_2(M)$  con el aumento de  $nff$  y los diagramas en el espacio de las coordenadas demoradas (Figuras 2 c) y d)). En la Figura 2 b) puede verse que para  $nff=2$  el comportamiento es el esperado para un ruido blanco, i.e., la ausencia de valor límite en la dimensión de correlación; mientras que al aumentar el valor del parámetro de filtrado (a pesar de que no es posible observar estructura a simple vista en las series temporales) la pendiente en el gráfico  $D_2(M)$  vs.  $M$  comienza a disminuir. A partir de  $nff=25$  se observa un valor límite finito lo cual indicaría una dimensión igualmente finita para el atractor; en el análisis de series temporales esto último se toma como indicador de dinámica no lineal subyacente en el sistema. Para el caso de estudio, se puede afirmar que es el resultado del filtro simple que aplicamos a la señal aleatoria original.

Si ahora tomamos en cuenta la posibilidad de que una señal asociada a la evolución de un sistema caótico con un vector de estado de  $n$  de coordenadas independientes, se encuentre superpuesta a un ruido blanco de fondo, el empleo de un filtro de similares características al utilizado en el presente trabajo, puede inducir a suponer un número incorrecto  $n'$  de grados de libertad activos.

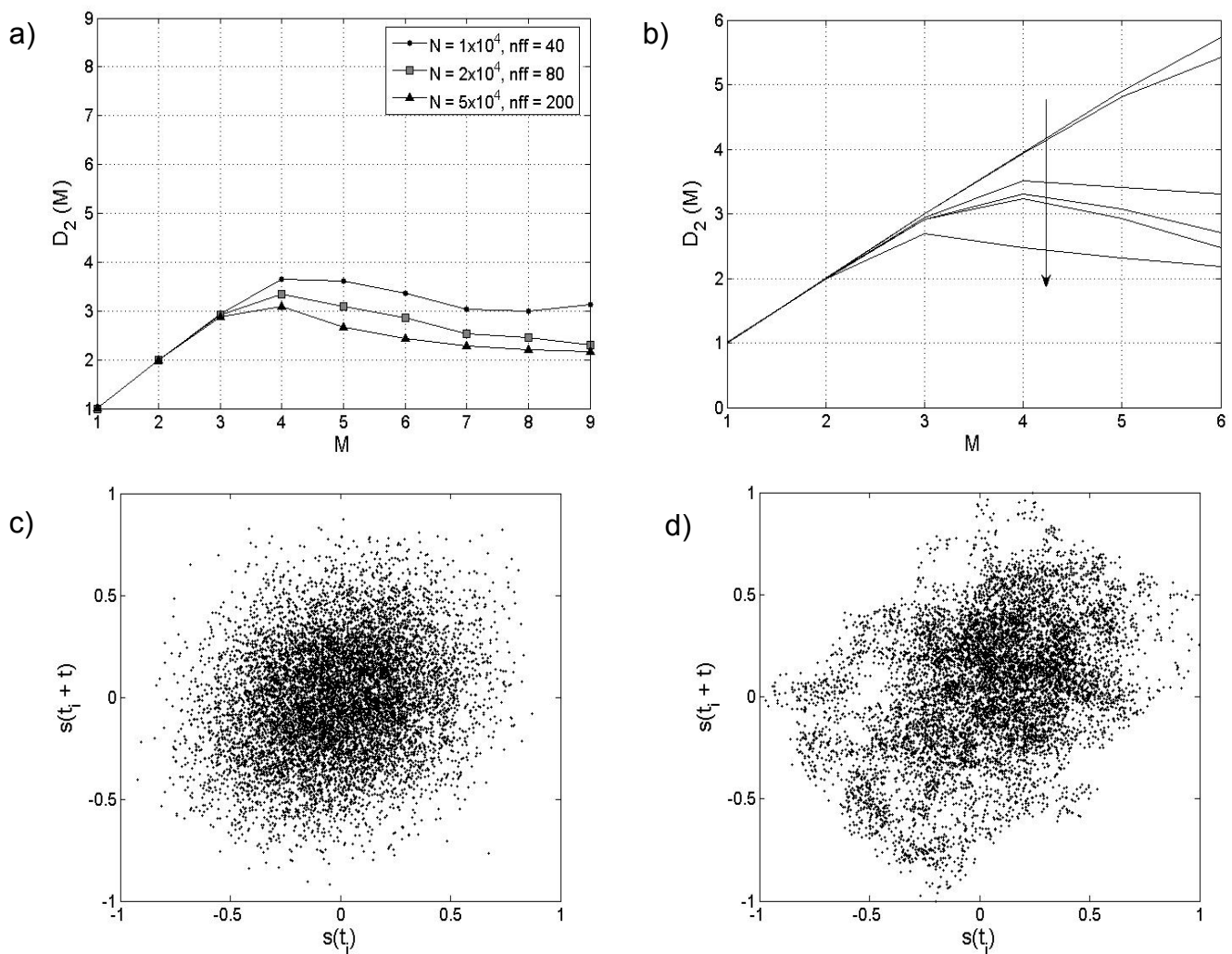


Fig. 2: Resultados de las simulaciones: a) dimensión de correlación para tres tamaños de series con el parámetro  $nff$ ; b) efecto del aumento de  $nff$  en la serie de valores sobre  $D_2(M)$ ; c) y d) representaciones de los dos valores extremos de  $nff$  del panel (b), 2 y 200 respectivamente.

## CONCLUSIONES

En el presente trabajo mostramos como la aplicación de un esquema de filtrado estándar en un conjunto de datos sin ningún tipo de estructura subyacente es capaz de producir el mismo resultado que un sistema con verdadera dinámica caótica según un análisis de detección de comportamiento no lineal ampliamente usado. El tamaño de la ventana utilizada en los promedios resulta crítico en el esquema de filtrado; si trasladamos este esquema de razonamiento al análisis de datos producidos por sistemas extendidos de reacción difusión (Rafti et al., 2007, Uecker 2007), los resultados son útiles para predecir los efectos esperables al utilizar un esquema de discretización demasiado grande para la escala temporal/espacial inherente al sistema. Esto es, la utilización de esquemas de discretización inadecuados puede dar como resultado la caracterización de sistemas sin estructura como falsos regímenes caóticos.

## AGRADECIMIENTOS

Los autores de este trabajo agradecen el apoyo de la UNLP (Universidad Nacional de La Plata), CICPBA (Comisión de Investigaciones Científicas de la Provincia de Buenos Aires) y CONICET (Consejo de Investigaciones Científicas y Tecnológicas), a través de los subsidios y becas otorgadas.

**REFERENCIAS**

- Akay, M., *Nonlinear Biomedical Signal Processing (vol. II)*, IEEE Press Series on Biomedical Engineering, Nueva York, EE.UU (2001).
- Argyris, J. y Andreadis, I., *The influence of Noise on the Correlation Dimension of Chaotic attractors*, Chaos, solitons and fractals: 9(3), pp. 343-361 (1998).
- Harikrishnan, K. P., Misra, R., Ambika, G. y Kembhavi, A.K., *A non-subjective approach to the GP algorithm for analysing noisy time series*, Phys. D: 215, pp. 137-145 (2006).
- Hongler, M.O., Filliger, R., Blanchard, Ph. y Rodriguez, J., *Noise Induced Temporal Patterns in Populations of Globally Coupled Oscillators*, 2nd International Workshop on Nonlinear Dynamics and Synchronization, Klagenfurt, Austria, 20 – 21 de Julio (2009).
- Imbihl, R. y Ertl, G., *Oscillatory kinetics in Heterogeneous Catalysis*, Chem. Rev., 95, 697-733 (1995).
- Kantz, H. y Schreiber, T. *Nonlinear Time Series Analysis (2da edición)*, Cambridge University Press, Cambridge, Inglaterra (2003).
- Makarenko A., *Different types of chaotic behavior for different space and time scales in complex systems*", 2nd International Workshop on Nonlinear Dynamics and Synchronization, Klagenfurt, Austria, 20 – 21 de Julio (2009).
- Osborne, A.R. y Provenzale, A., *Finite Correlation Dimension for Stochastic Systems with Power-Law Spectra*, Physica D: 35, pp. 357-381 (1989).
- Rafti, M., Lovis, F., Zeng, Y. e Imbihl, R., *Homogeneous and front-induced surface transformations during catalytic oxidation of ammonia over Pt(100)*, Chem. Phys. Lett.: 446, pp. 323-328 (2007).
- Rafti, M., Vicente, J.L., Uecker, H. e Imbihl, R., *Simulation of the life cycle of adsorbate islands on the Pt(100) surface during the NO + NH<sub>3</sub> reaction*, Chem. Phys. Lett.: 421, pp. 577-583 (2006).
- Rosso, O. A., y otros autores, *Distinguishing Noise from Chaos*, Phys. Rev. Lett.: 99, pp. 154102 1-4 (2007).
- Ramdani, S., Bouchara, F. y Casties, F., *Detecting determinism in short time series using a quantified averaged nearest neighbors approach*, Phys. Rev. E: 76, pp. 036204 1-14 (2007).
- Packard, N.H., Crutchfield, J.P., Shaw R.D. y Farmer F.D. *Geometry from a time series*, Phys. Rev. Lett.: 45, pp. 712-716 (1980).
- Uecker, H., *Pattern formation for NO+NH<sub>3</sub> on Pt(100): Two-dimensional numerical results*, Phys. Rev. E: 71, pp. 016207 (2005).
- Zhdanov, V.P., *Comment on "Modelling pattern formation in CO+O<sub>2</sub> (reaction) on Pt{100}"*, Chem. Phys. Lett.: 390, pp. 301-304 (2004).