

Análisis de datos caóticos

Chorro, J.L. y Lopez, A.
Universidad de Valencia y Universidad de Sevilla

La finalidad de este trabajo es la de comparar tres procedimientos de filtrado de datos de series temporales cuyo objeto es el de paliar el efecto perturbador del ruido aleatorio en los resultados de las técnicas de análisis de datos caóticos. Se ha aplicado los métodos de suavización, descomposición matricial singular y predicción no lineal a series temporales generadas mediante la ecuación logística y a las que se ha añadido error aleatorio. Seguidamente se ha comparado las estimaciones de los exponentes de Lyapunov, dimensiones de correlación y capacidad, espectro de potencia y otras bajo dichas condiciones.

The main goal of this paper is to compare three procedures of data filtering in order to alleviate the effect of the noise on existing techniques of chaotic data analysis. The methods of data smoothing, singular value decomposition and nonlinear prediction have been applied to logistic time series in which some amount of error has been added. The estimates of Lyapunov coefficients, correlation and capacity dimension, power spectra, and others have been compared under such conditions.

La finalidad de este trabajo es la de comparar tres procedimientos de filtrado de datos de series temporales cuyo objeto es el de paliar el efecto perturbador del ruido aleatorio en los resultados de las técnicas de análisis de datos caóticos. Durante los últimos años se ha venido proponiendo diversas técnicas de análisis de datos cuya finalidad es la de detectar caos determinista (Abarbanel, 1996, Ott, Sauer y Yorke, 1994). La variación caótica presenta características parcialmente similares a las del ruido aleatorio, y ello plantea la dificultad de que técnicas sensibles al caos determinista también son afectadas por el ruido aleatorio y pueden inducir a conclusiones erróneas (Chon et al., 2000), como son la obtención de coeficientes de Lyapunov positivos o la obtención de dimensiones finitas. Este problema ha sido abordado desde enfoques diversos, como son el desarrollo de métodos robustos de análisis de datos o el filtrado previo de los datos con la finalidad de minimizar la presencia de variación aleatoria. Un enfoque consiste en obtener predicciones que se asume resultan libres de error, o al menos contienen cantidades sustancialmente menores de error que los datos originales. Las técnicas de predicción que se ha propuesto han consistido generalmente en aplicar procedimientos ya conocidos a segmentos de las series temporales lo suficientemente pequeños como para hallarse dentro del intervalo de predecibilidad o para exhibir características de linealidad. Casdagli (1989) propone una técnica consistente en localizar patrones históricos similares a corto plazo al del momento en el que interesa la predicción. Las predicciones consisten en promedios de lo que sucedió a continuación. En otros casos, las predicciones se han basado en ajustes polinómicos locales a los vecinos más próximos (Casdagli, op.cit.). Otra técnica ha sido propuesta por Sugihara (1990) y consiste en obtener pesos que representen el distanciamiento de los vecinos más próximos de valores de segmentos locales de la serie. Otros procedimientos se basan en la serie temporal completa, como es el procedimiento de descomposición singular de la matriz de intercorrelaciones de los retardos de la serie, en que se pretende descomponer la variación en un conjunto de componentes ortogonales que representen las principales fuentes de variación. A

continuación se sustituye la serie original por la obtenida a partir del modelo. En cuanto a los resultados, Abarbanel (op.cit.) señala que no han sido tan satisfactorios como se esperaba, y además que, en la práctica, procedimientos sencillos a veces han dado mejor resultado que otros más sofisticados. A fin de estudiar el efecto de la variación aleatoria en los resultados de algunos de los procedimientos actualmente en uso para la detección del caos determinista se ha generado varias series temporales mediante ordenador, a las que se ha añadido error aleatorio. Posteriormente se ha aplicado tres procedimientos de filtrado, el basado en la suavización de los datos originales, el basado en la descomposición singular de la matriz y el basado en predicción no lineal. El procedimiento de suavización consiste en sustituir cada dato por el promedio de él mismo y los dos datos adyacentes. El método de descomposición singular consiste en sustituir los datos de la serie original por los obtenidos a partir de los componentes principales que la representan. Seguidamente se ha aplicado varias técnicas de análisis de datos caóticos y se ha comparado los resultados a fin de estudiar qué procedimiento es más adecuado.

Método

Se ha generado una serie temporal de 4000 datos mediante la ecuación logística:

$$Y_{t+1} = 4Y_t(1 - Y_t)$$

Se ha obtenido el diagrama de fase, diagrama de retorno, prueba de agrupamiento, espectro de potencia, exponente de Lyapunov, dimensión de capacidad, dimensión de correlación, correlograma, descomposición matricial por valores singulares, coeficiente de entropía y error de predicción. Seguidamente se ha generado otra serie de idéntico tamaño con la siguiente expresión:

$$Y_{e,t} = Y_t + e_t \cdot 0,1$$

donde los valores de e_t son seleccionados aleatoriamente de una distribución normal con media 0 y varianza igual a 1. Se ha repetido en esta serie los procedimientos citados y a continuación se ha obtenido nuevas series mediante filtrado de la anterior, repitiéndose en cada una de ellas la aplicación de los procedimientos de detección de caos determinista. Además se ha generado una serie temporal que consta de solo ruido (los valores han sido obtenidos de una distribución uniforme

Correspondencia: Juan Luis Chorro Gascó
Departamento de Metodología. Facultad de Psicología.
Universidad de Valencia
Av. Blasco Ibañez 21 46101 Valencia
E-mail: juan.l.chorro@uv.es
Investigación financiada con cargo al proyecto BSO2001-2904 del Ministerio de Ciencia y Tecnología.

en el intervalo 0 - 1) a fin de comparar los anteriores resultados con los obtenidos en presencia de solo error.

Resultados

La comparación de los resultados obtenidos en la serie a la que no se ha añadido error aleatorio (serie de referencia) con la serie a la que sí se ha añadido dicho error muestra, en primer lugar, una distorsión de la imagen del diagrama de fase, que aparece completamente nítido en la serie original (Fig.1) y en cambio muestra un aspecto nebuloso en la serie a la que se ha añadido error (Fig. 2). La gráfica de la prueba de agrupamiento es menos nítida en la serie con error (Fig.4), pero se asemeja bastante a la gráfica de referencia (Fig. 3). Las gráficas del espectro de potencia son diferentes en ambos casos (Figs.5 y 6). En cuanto a la determinación de la dimensión del atractor, la dimensión de capacidad estimada con la serie a la que se ha añadido error es igual a 4.109, observándose estabilización en D=5 (Fig.7), en tanto que la dimensión estimada en la serie original es de 1.196, observándose estabilización a partir de D=2 (Fig.8). La dimensión de correlación estimada a partir de la serie original es igual a 0.95, estable a partir de D=1, en tanto que en la serie con error es igual a 4.109, estabilizándose a partir de D=5 (Figuras 9 y 10). Las funciones de autocorrelación presentan perfiles muy análogos (Figuras 11 y 12), y lo mismo sucede con el número de autovalores de la matriz de correlaciones. El perfil de la gráfica del error de predicción promedio acusa la presencia del error aleatorio pasando de un aspecto sigmoidal al de curvatura cóncava (Figuras 13 y 14).

Comparación con los resultados obtenidos mediante triple filtrado:

El diagrama de fase (Fig.15) presenta un aspecto nebuloso, muy poco parecido a las gráficas de la serie original. La gráfica de la prueba de agrupamiento muestra claramente la presencia de estructura en los datos (Fig.16). El espectro de potencia aparece truncado para las frecuencias más altas (Fig.17), y la estimación de la dimensión de capacidad es igual a 2.074, con estabilización a partir de D=2 (Fig.18), y la dimensión de correlación es 3.682, con estabilización imperfecta a partir de D=5 (Fig.19). La gráfica de la función de autocorrelación presenta una caída más suave que la de la serie original (Fig.20), y lo mismo sucede con la de los autovalores de la matriz de correlación. El índice de complejidad relativa LZ es igual a .83, el mayor coeficiente de Lyapunov es igual a 0.89 y el coef. de entropía es igual a 0.509.

Filtrado por descomposición singular matricial:

El diagrama de fase (Fig.21) y la gráfica de la prueba de agrupamiento (Fig.22) presentan un aspecto nebuloso en el que no se distingue estructura subyacente. El espectro de potencia (Fig.23) presenta dos zonas de mayor energía, lo que la diferencia de la

distribución del espectro de la serie original. La gráfica de la dimensión de capacidad muestra estabilización a partir de D=5 (Fig.24) y la de dimensión de correlación no muestra estabilización (Fig.25). El índice de complejidad relativa LZ es igual a .91, el mayor coeficiente de Lyapunov es igual a 0.808 y el coef. de entropía es igual a 0.528.

Filtrado mediante predicción no lineal:

La gráfica de agrupamiento (Fig.26) muestra un aspecto nebuloso, con un leve indicio de presencia de estructura. Lo mismo sucede con la gráfica de fase (Fig.27). El espectro de potencia muestra amplia distribución, si bien con preponderancia de los componentes de alta frecuencia (Fig.28). La estimación de la dimensión de capacidad es de 3.3, ocurriendo la estabilización a partir de D=5 (Fig.29). La dimensión de correlación estimada es de 3.827, notándose tendencia a la estabilización a partir de una dimensión de embebido igual a 4 (Fig.30). La función de correlación muestra un pico inicial negativo, siendo similar a la de la serie de referencia. El índice LZ es igual a 1.03, el coeficiente de Lyapunov es igual a 1.036 y el coeficiente de entropía es igual a 0.59.

A modo de resumen, la siguiente tabla facilita la comparación de los índices obtenidos:

Error	Filtrado	λ	LZ	Dc	Dr	r
Aleatorio	.9	1.04	2.3	4.4	.64	
0	No	.95	1.03	1.2	1.87	.62
10	No	.95	1.06	2.8	4.11	.63
10	Suavizado	.89	0.83	2.52	3.85	1.91
10	Desc.Sing	.81	.91	2.86	4.26	.61
10	Pred.NLin	1.04	1.03	3.31	3.83	.53

(La primera fila muestra los resultados obtenidos de los análisis de datos realizados con la serie que contiene solo ruido. La segunda fila muestra los resultados obtenidos con la serie en que no ha añadido error aleatorio. La tercera fila muestra los resultados obtenidos con la serie a la que se ha añadido error aleatorio sin aplicar ninguna de las técnicas de filtrado. Las restantes filas muestran los resultados obtenidos con cada una de las técnicas de filtrado).

Conclusiones

Se observa que el procedimiento de filtrado de datos que permite una mejor estimación de la dimensión del atractor (tanto de capacidad como de correlación) es el triple suavizado. Por contra, es el que da lugar al valor de autocorrelación más alejado del valor de referencia, así como con el exponente de Lyapunov. A modo de conclusión final parece más recomendable el procedimiento de triple filtrado por ser el que mejor aproxima la estimación de la dimensión del atractor.

Figuras

Figura 1.

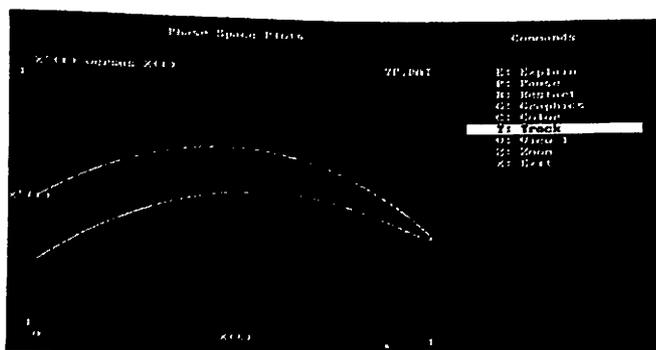


Figura 2.

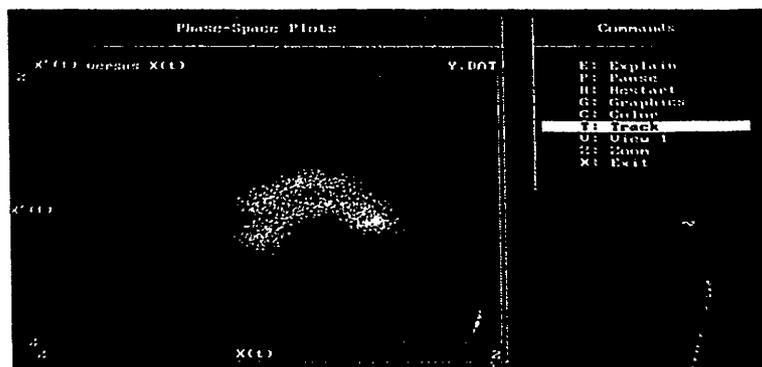


Figura 3.



Figura 4.

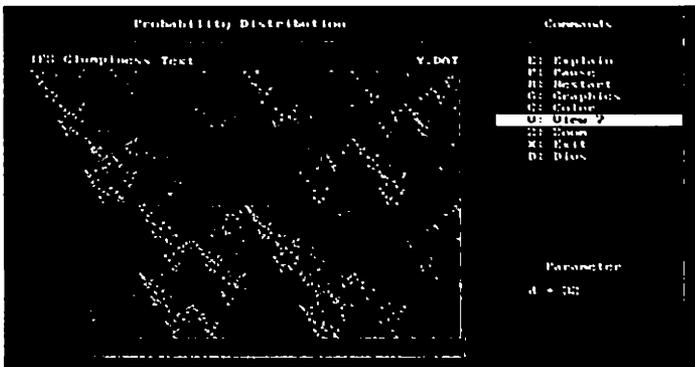


Figura 5.

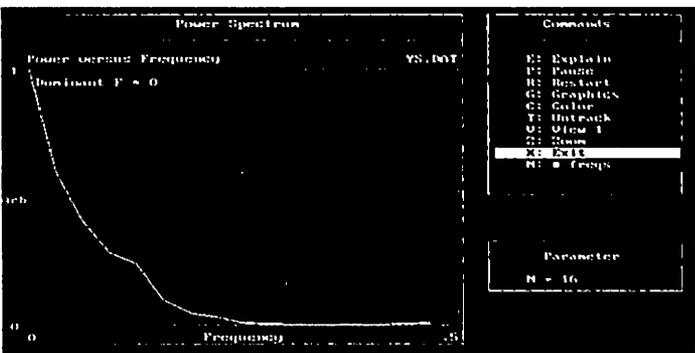


Figura 6.

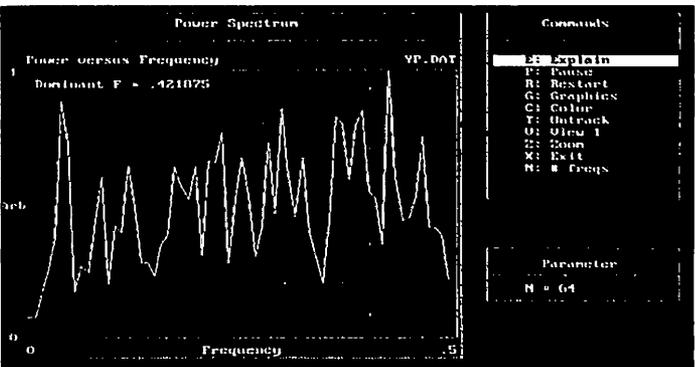


Figura 7.

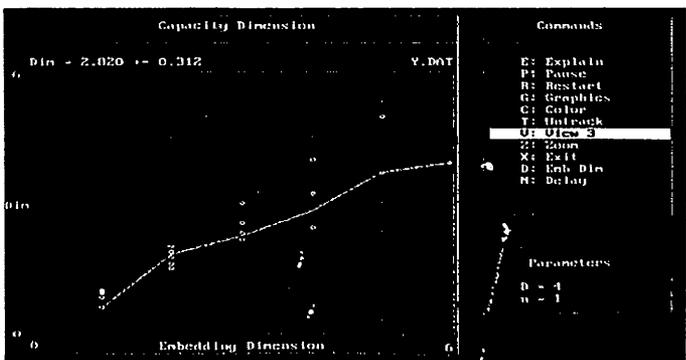


Figura 8.

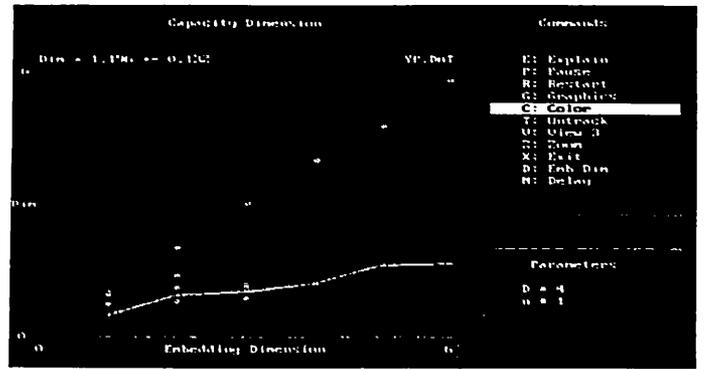


Figura 9.

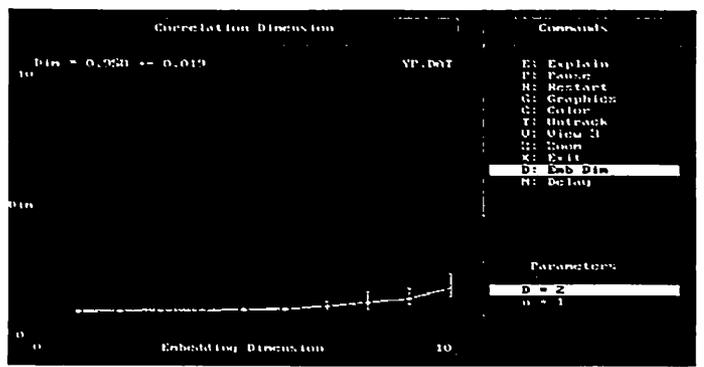


Figura 10.

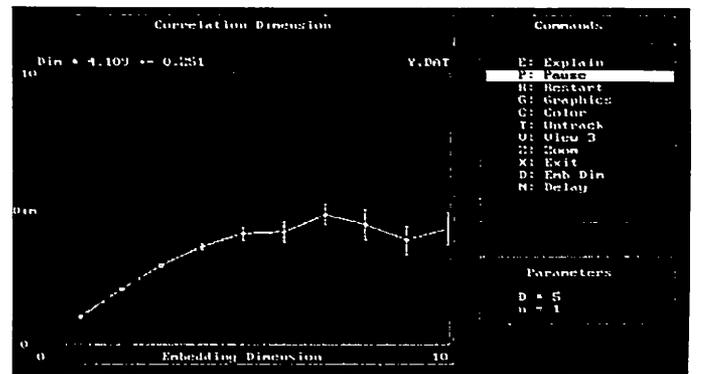


Figura 11.

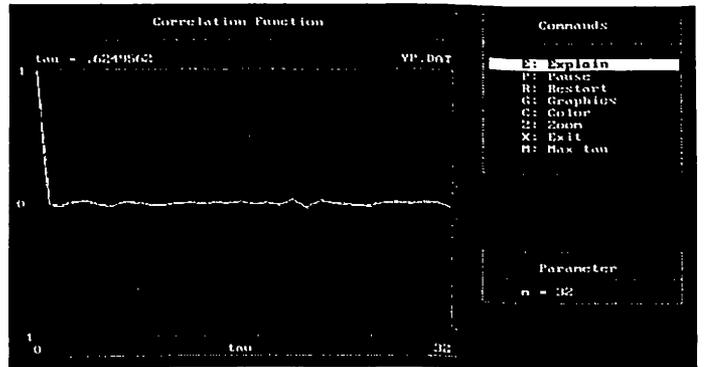


Figura 12.

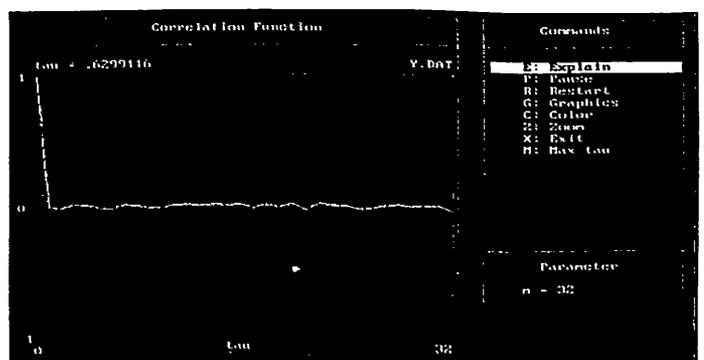


Figura 13.

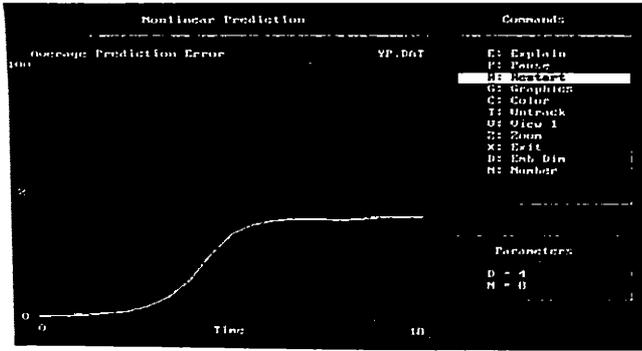


Figura 18.

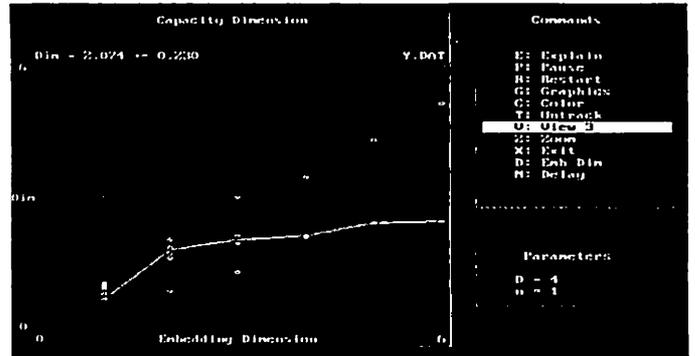


Figura 14.

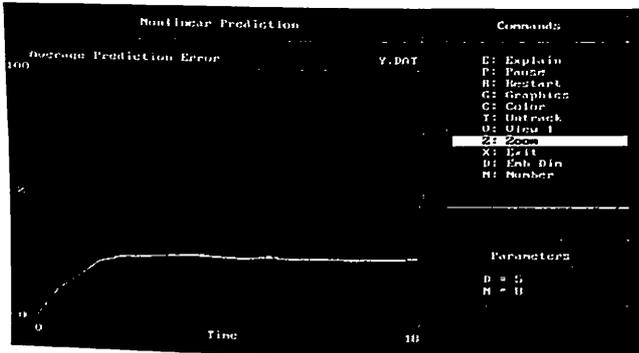


Figura 19.

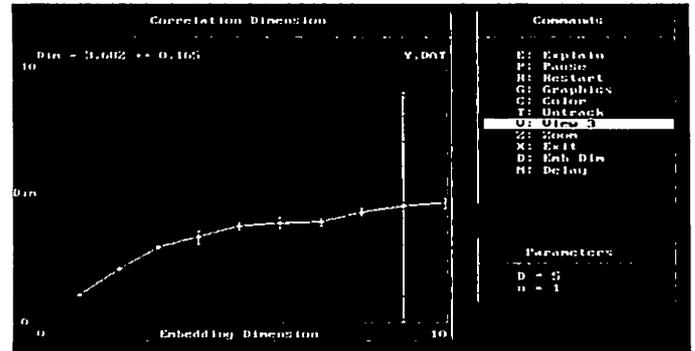


Figura 15.

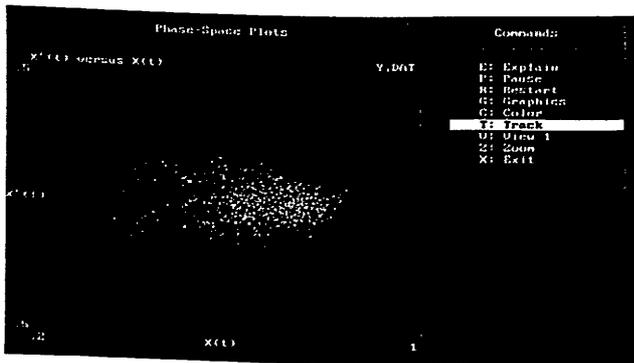


Figura 20.

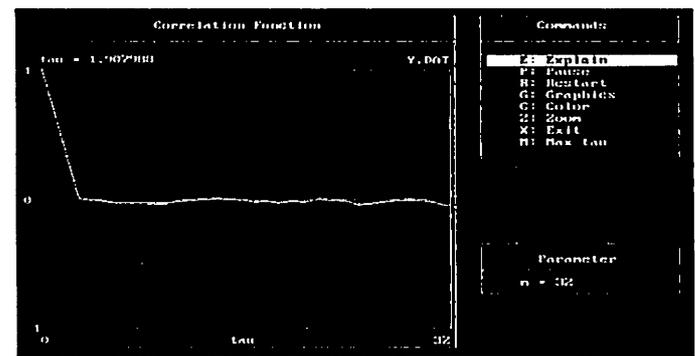


Figura 16.

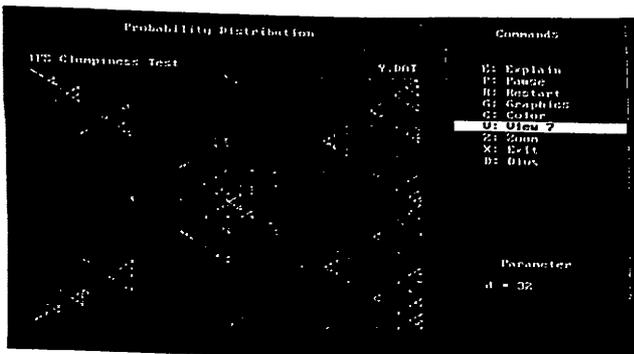


Figura 21.

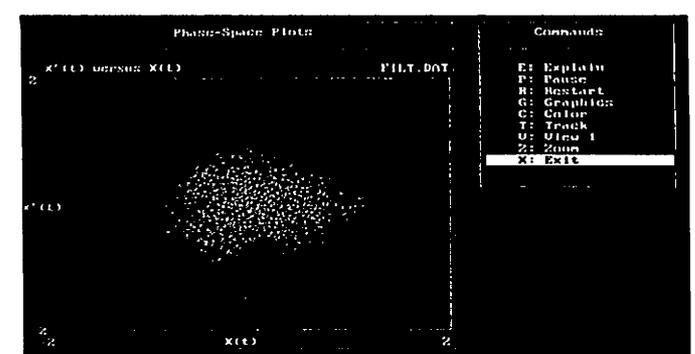


Figura 17.

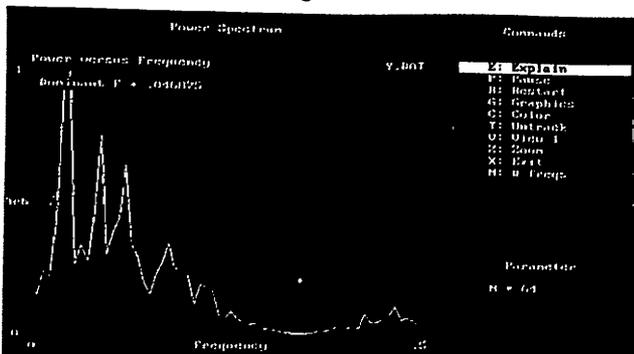


Figura 22.

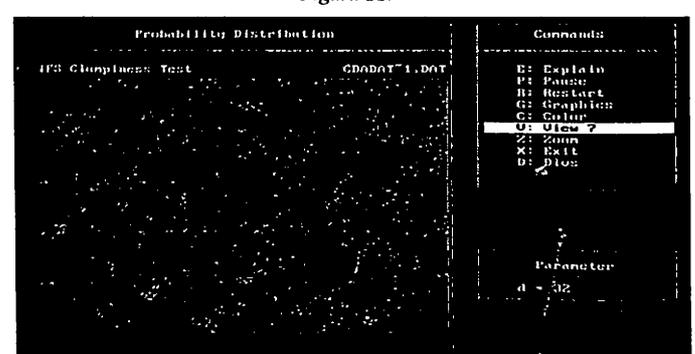


Figura 23.

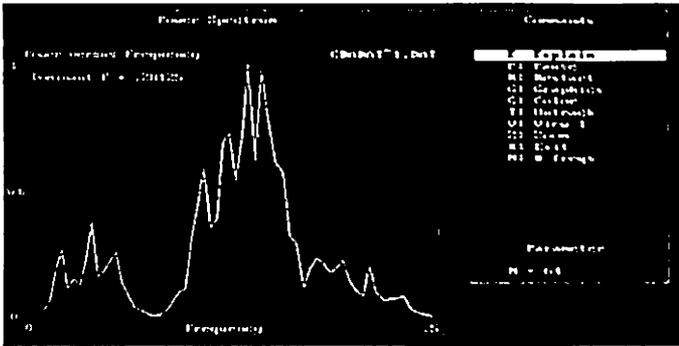


Figura 24.

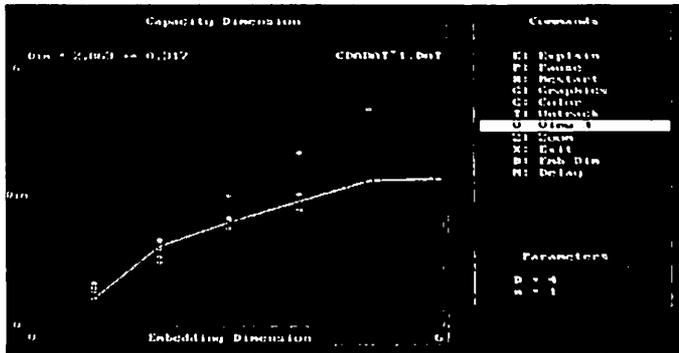


Figura 25.

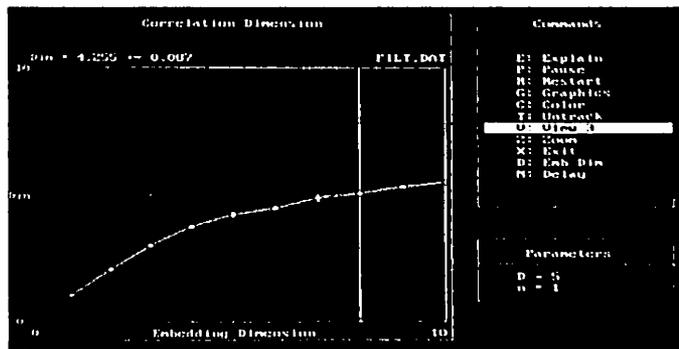


Figura 26.

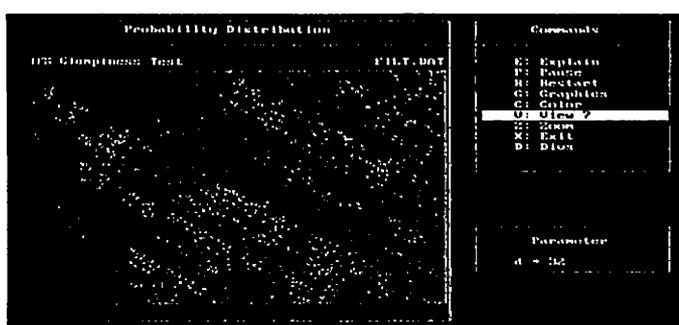


Figura 27.

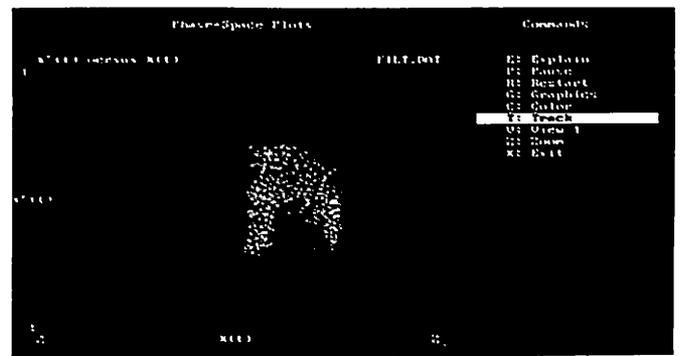


Figura 28.

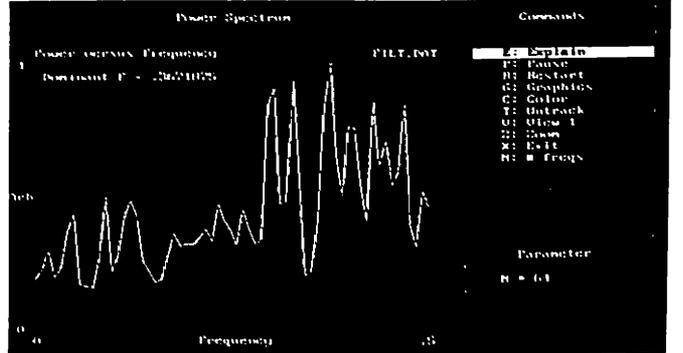


Figura 29.

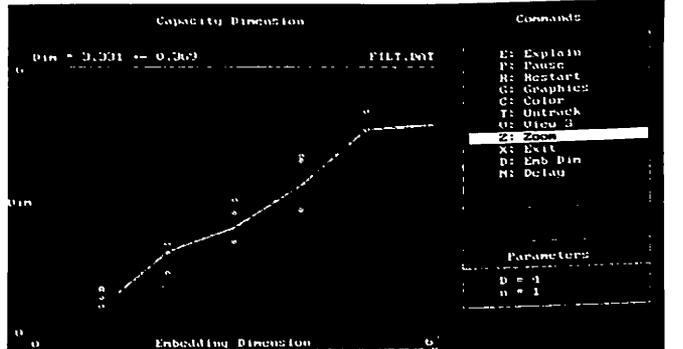
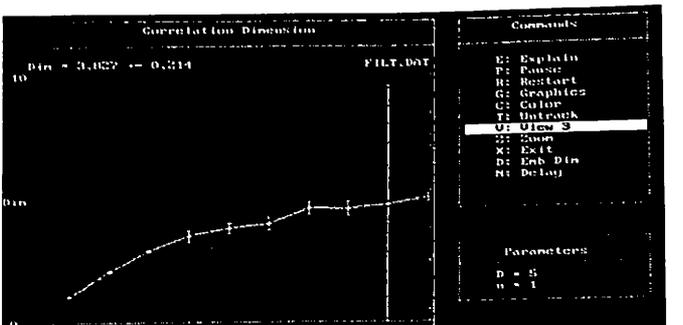


Figura 30.



Referencias

- Abarbanel,H.D. (1996) *Analysis of observed Chaotic Data*. New York, Springer.
- Abraham,F.D., Gilgen,A.R.(1995) *Chaos Theory in Psychology*. Westport, Greenwood.
- Brown,C. (1995) *Chaos and Catastrophe Teories: Nonlinear modeling in the Social Sciences*. Thousand Oaks, Sage Pub.
- Casdagli,M. (1989) Nonlinear prediction of chaotic time series. *Physica, D35*.
- Chon,K.H., Yip,K.P., CaminoB.M., Marsh,D.J. y Holstein-Rathlou,N. (2000) Modeling non linear determinism in short time series from noise driven discrete and continuous systems, *International Journal of Bifurcation and Chaos*, 10(12), 2745-2766.
- Cambel,A.B.(1992) *Applied Chaos Theory: A Paradigm for Complexity*. New York, Academic.
- Grassberger,P., Hegger,,R, Kantz,H. Y Shaffrath,C. (1994) "On noise reduction methods for chaotic data" en Ottt,E., Sauer,T. y Yorke,J.A. (Eds.) *Coping with chaos*. New York, Wiley.
- Ott,E., Sauer,T. y Yorke,J.A. (1994) *Coping with Chaos: Analysis of Chaotic Data and the exploitation of chaotic systems*- New York, Wiley.
- Robertson,R., Combs,A. (1995) *Chaos Theory in Psychology and the Life Sciences* . L.Erlbaum.
- Sprott,J.C. y Rowlands,G. (1995) *Chaos Data Analyzer*. New York, American Institute of Physics.
- Sugihara,G. Y May,R.M. (1990) "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series". *Nature*, 344, 734-740.
- Tufillaro,N.B. y Abbot,T. (1992) *An Experimental Approach to Nonlinear Dynamics and Chaos*. New York, Addison-Wesley.