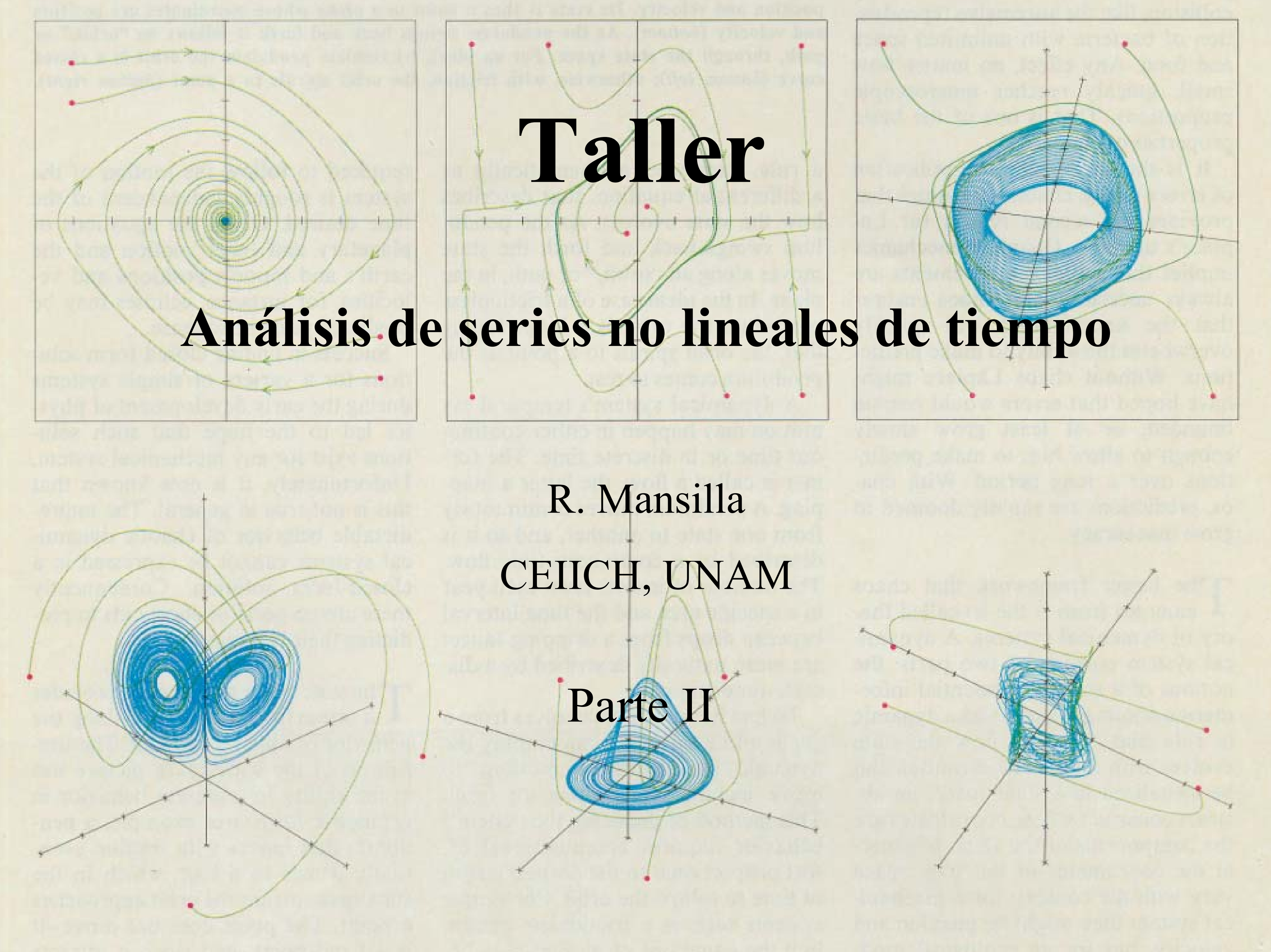


Taller

Análisis de series no lineales de tiempo

R. Mansilla

CEIICH, UNAM

Parte II

La función de información mutua es una medida de correlación más potente que las tradicionales medidas estadísticas.

C. J. Celluci, *et al.*, “Statistical validation of mutual information calculations: comparison of alternative numerical methods”

Physical Review E, vol. **71**, 066208, 2005.

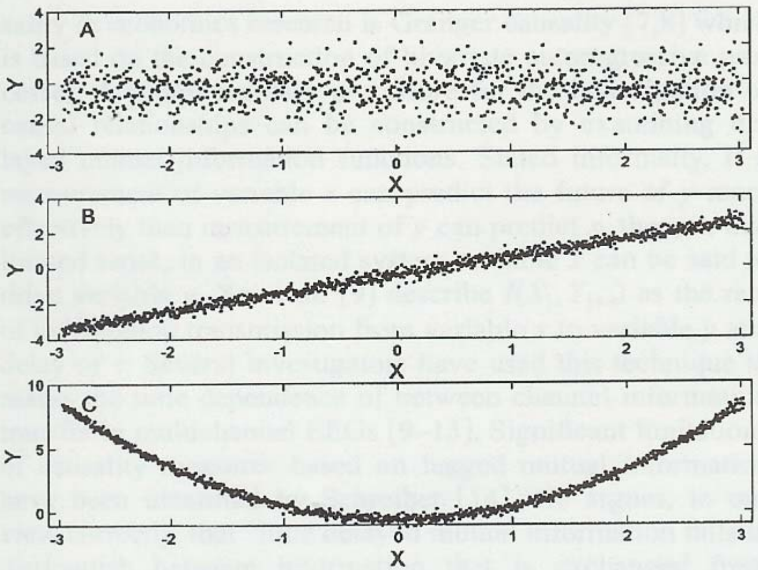
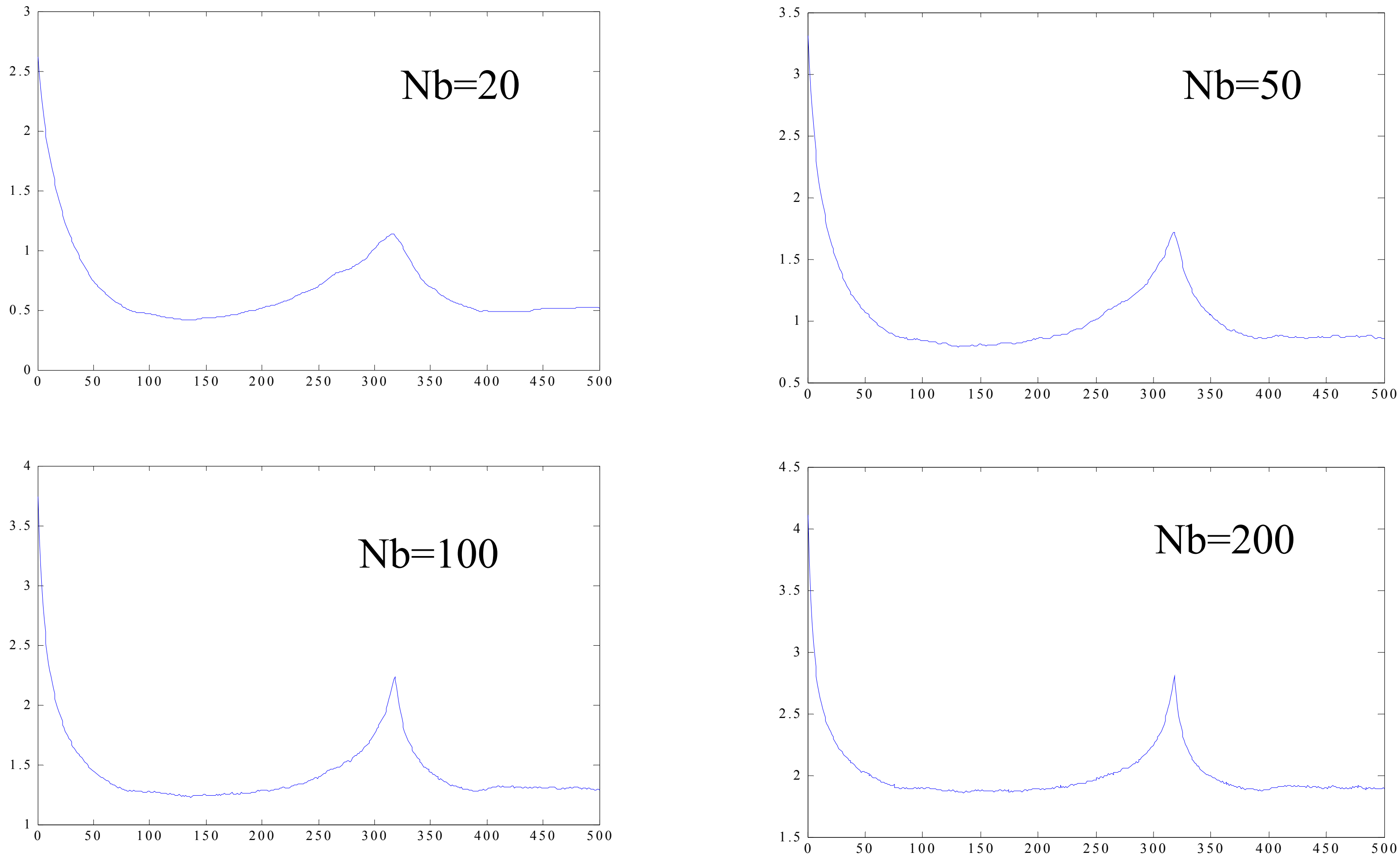


FIG. 1. Data sets used in the correlation study of Table I. In each case, x varies from −3 to +3 in steps of 0.0006. (A) $y_i=\varepsilon_i$, a normally distributed random variable with zero mean and unit variance. (B) $y_i=x_i+0.2\varepsilon_i$. (C) $y_i=x_i^2+0.2\varepsilon_i$.

TABLE I. Correlation analysis.

	Pearson r	Pearson P_{null}	Spearman r_S	Spearman P_{null}	Kendall's tau	Kendall's P_{null}	$I(X,Y)$	$I(X,Y)P_{null}$
Normally distributed random	−0.0037	0.7112	−0.0040	0.6854	0.0027	0.6845	0.1356	0.7851
Linearly correlated	0.9934	0	0.9936	0	0.9270	0	2.9186	0
Parabolically correlated	0.0001	0.9912	$<10^{-4}$	0.9928	$<10^{-5}$	0.9989	3.0304	0

No obstante, para su cálculo se debe tener una estimación muy precisa de la función de densidad conjunta $p(x,y)$ que se estima a través de $P_{X,Y}(i,j)$.



Rosslerx.txt

¿Cuál es el tamaño óptimo del histograma?

F. Mosteller, J. W. Tukey, “Data Analysis and Regresion” Addison-Wesley pag. 49, 1977.

$$Nb = \sqrt{n}$$

J. B. Bendat, A. G. Piersol, “Measurement and Analysis of Random Data” John Wiley, pag. 284, 1966.

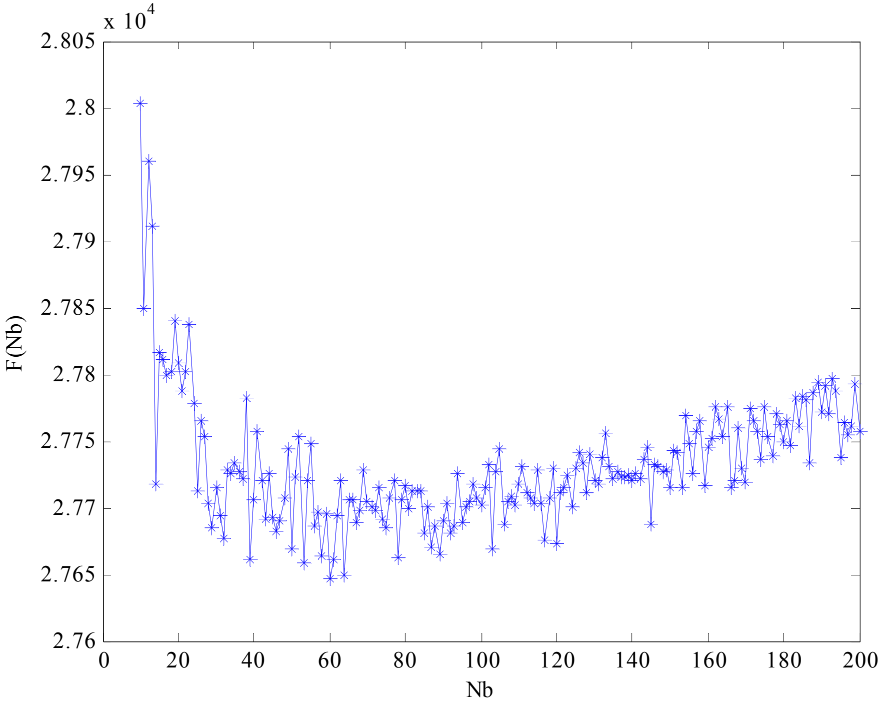
$$Nb = 1.87(n-1)^{0.4}$$

Y. Rissanen, “Stochastic Complexity in Statistical Inquiry, World Scientific pag. 76, 1992.

$$F(Nb) = n\log_2\left(\frac{R}{Nb}\right) + \log_2\binom{n}{n_1,\ldots,n_{Nb}} + \log_2\binom{n+Nb-1}{n}$$

La mínima longitud de descripción se alcanza para el valor de Nb donde se minimiza la complejidad estocástica F

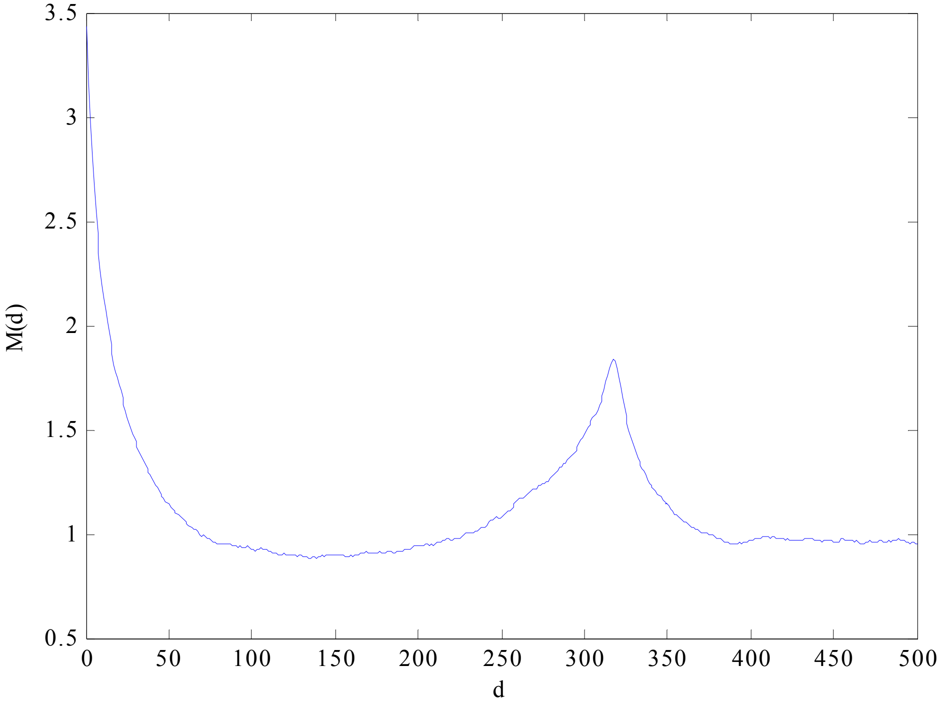
$R = x_{\max} - x_{\min}$ y $n_1, \ldots, n_{Nb}$ ← Son las diferentes ocupaciones de un histograma de Nb elementos.



El valor mínimo de la función de Rissanen se alcanza en $Nb = 60$.

El valor mínimo de la función de información mutua se alcanza en:

$$d = 135$$



De acuerdo a lo que hemos visto con anterioridad:

$$\tau = 135$$

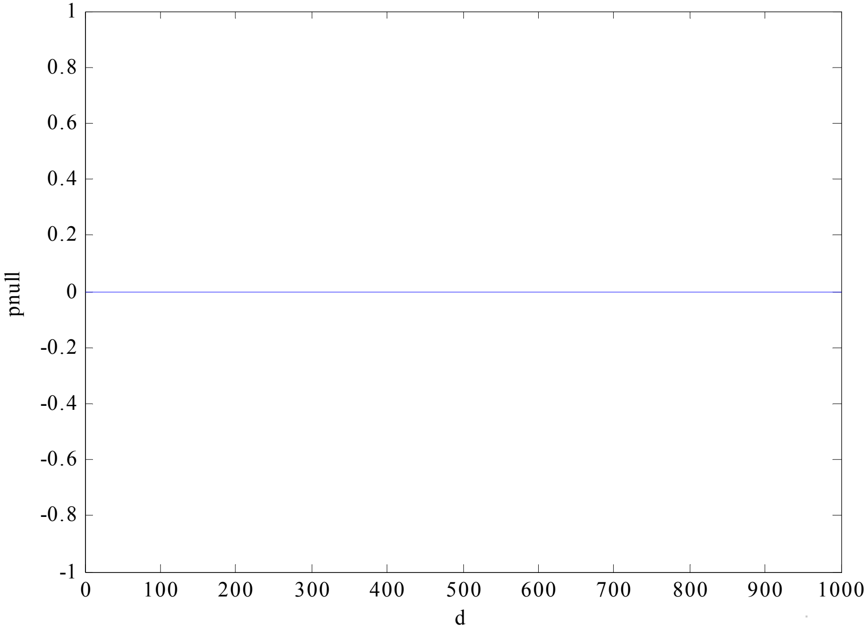
La utilización del primer mínimo de la función de información mutua como el retraso óptimo no garantiza que en esa distancia la información mutua sea cero. Por tanto debemos tener un criterio estadístico que nos permita rechazar o aceptar la hipótesis nula de la independencia estadística:

$$X,Y \quad independientes \quad \Rightarrow \quad I(X,Y)=0$$

$$\chi^2 = \frac{1}{n}\sum_{i=1}^{Nb}\sum_{j=1}^{Nb}\frac{\left[n\theta(i,j)-n_i n_j\right]^2}{n_i n_j} \qquad \nu = (Nb-1)^2$$

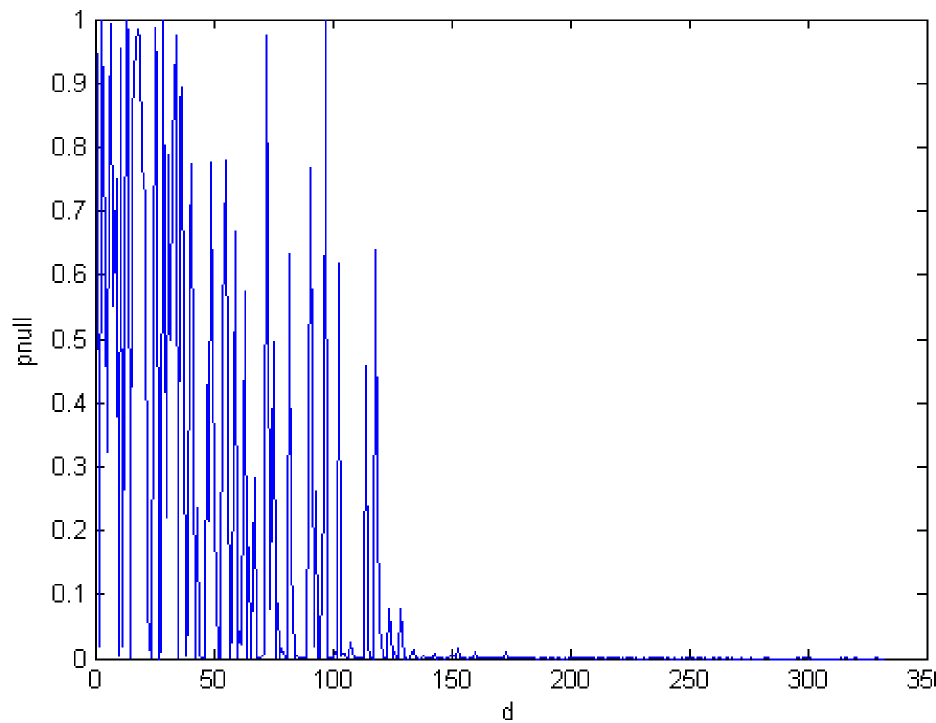
Con esto, la probabilidad de ocurrencia de la independencia estadística es:

$$p_{null} = Q\left(\frac{\chi^2}{2},\frac{\nu}{2}\right) = 1-\frac{1}{\Gamma\left(\frac{\chi^2}{2}\right)}\int_0^{\frac{\nu}{2}} e^{-t}t^{\frac{\chi^2}{2}-1}dt \qquad P_{X,Y}(i,j)=\frac{\theta(i,j)}{n}$$



Con los datos rosslerx.txt la probabilidad de independencia estadística es cero para todas las distancias calculadas.

¡ Con 10000 datos construidos con el generador de números aleatorios de MATLAB tenemos que para distancias mayores que 150 la probabilidad de independencia es practicamente cero !



Busqueda de la dimensión adecuada

En la busqueda de la dimensión de inmersión adecuada solo existen criterios heurísticos. El más usado esta relacionado con la llamada integral de correlación:

$$C_n(\varepsilon) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} H\left(\left\| X_i - X_j \right\| < \varepsilon\right)$$

Con esto, la dimensión de correlación se define como:

$$d_c = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\log C_n(\varepsilon)}{\log(\varepsilon)}$$

Basicamente, la dimensión de correlación es la pendiente de la integral de correlación en el origen.

El criterio para seleccionar una dimensión de inmersión es como sigue:

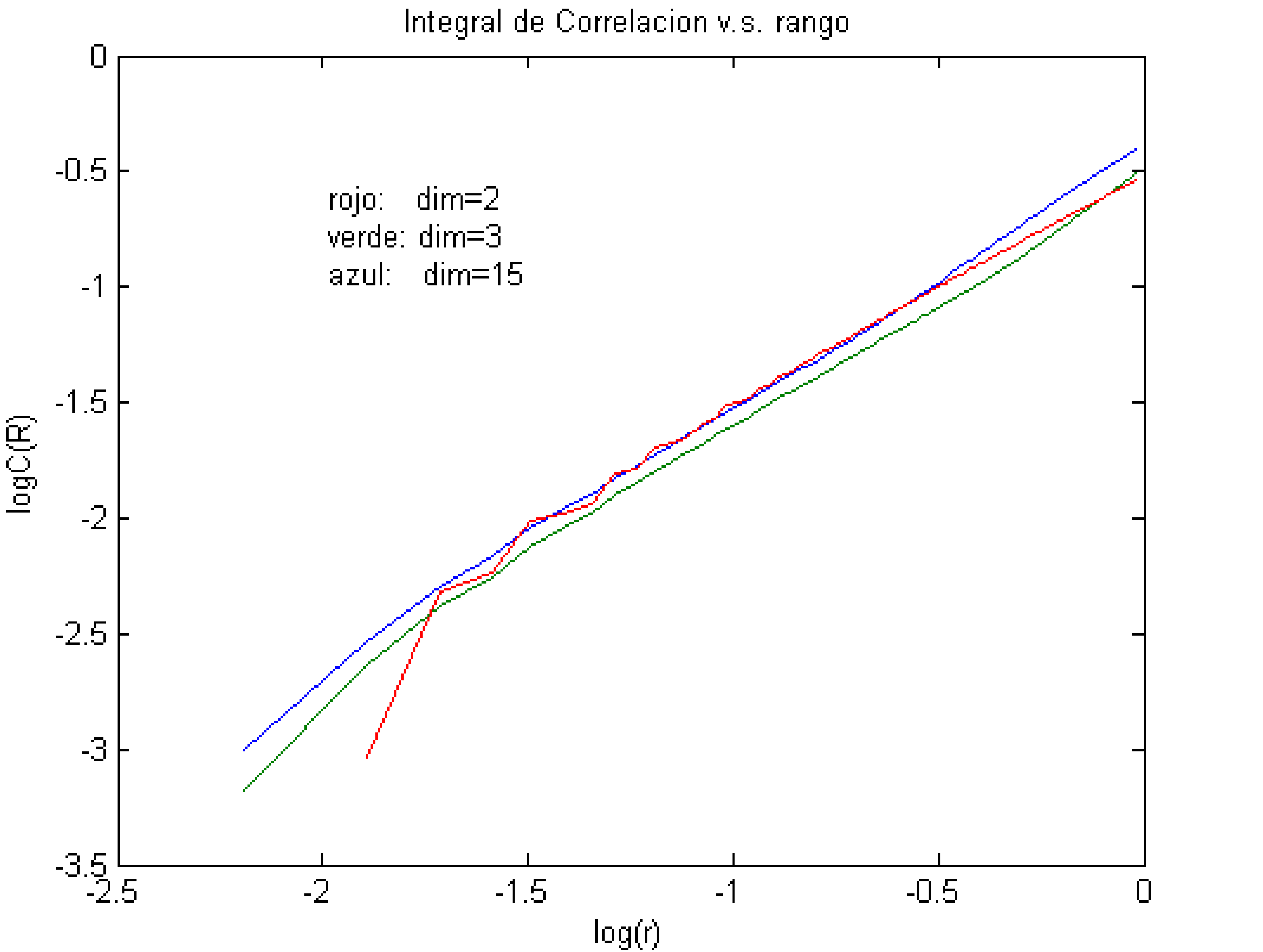
"*Elegir dimensiones de inmersión crecientes y en cada caso, calcular la integral de correlación. Cuando no se observen cambios en el comportamiento de la integral de correlación con respecto al incremento de la dimensión de inmersión, entonces se habrá encontrado un dimensión de inmersión adecuada*".

M. Small, "*Applied nonlinear time series analysis*", World Scientific, pag. 8, 2005.

Grassberger P., Procaccia I., Characterization of strange attractors. Physical Review Letters vol. **50**, pags. 346-349, 1983.

Theiler J., Spurious dimension from correlation algorithms applied to limited time-series data. Physical Review A vol. **34** pags. 2427-2432, 1986.

Albano A. M., Muench J., Schwartz C., Mees A. I., Rapp P. E., Singular-value decomposition and the Grassberger-Procaccia algorithm, Physical Review A vol. **38**: pags. 3017-3026, 1988.

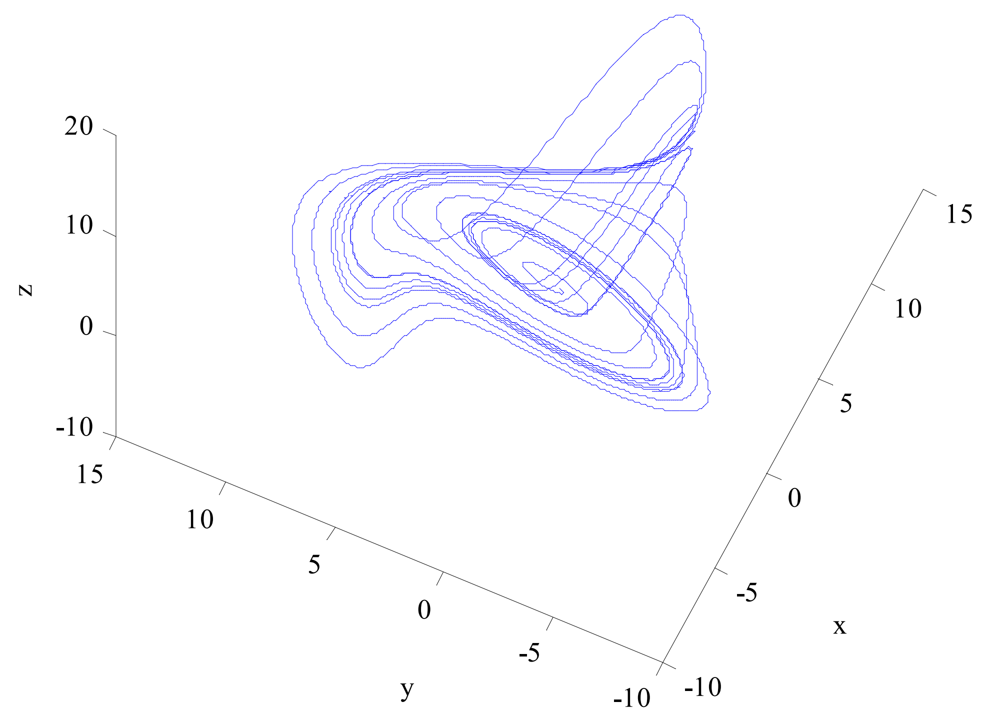


Integral de Correlacion v.s. rango
rojo: dim=2
verde: dim=3
azul: dim=15
logC(R)
log(r)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-2.5
-2
-1.5
-1
-0.5
0

Utilizando el programa **reconstruct1.m** para un valor del retraso igual a 135 y una dimensión de inmersión igual a 3 se obtiene la reconstrucción que se muestra abajo.

$\dim = 3$

$\tau = 135$



Fin