

# Comparación de métodos no lineales en la predicción de series temporales ruidosas

P.M. Granitto, P.F. Verdes, H.D. Navone y H.A. Ceccatto

*Instituto de Física Rosario, CONICET-UNR*

*Bv. 27 de Febrero 210 bis - (2000) Rosario - Argentina*

*e-mail: granitto@ifir.edu.ar*

Se discute la performance de métodos globales (redes neuronales) y locales (aproximación de hiperplano local) en la predicción de series temporales caóticas afectadas por ruido. Se investiga el deterioro de la performance de estos métodos en función de la longitud y la relación señal-ruido del registro histórico. Usando el mapa logístico y la ecuación de Mackey-Glass como ejemplos, se concluye que los métodos locales, que habían demostrado ser competitivos ante las redes neuronales para series puras<sup>1</sup>, resultan mucho más sensibles a la presencia de ruido.

We discuss the capabilities of global (neural network) and local (local hiperplane approximation) methods for the forecasting of chaotic time series with additive noise. We investigate the performance of these methods as a function of the database length and the signal-to-noise ratio. Using the logistic map and the Mackey-Glass equation as examples, we conclude that the local methods, which had been shown to be competitive with neural networks for clean series<sup>1</sup>, are more sensitive to noise.

## I. INTRODUCCIÓN

Se han propuesto muchos algoritmos para reconstruir la dinámica subyacente en una serie temporal y permitir predecir sus valores en un futuro próximo<sup>2,3</sup>. Dadas las observaciones  $\{x_t\}_0^N$  sobre un sistema, el problema consiste en la reconstrucción de la ley  $x_t = F(\mathbf{X}_t)$ , donde  $\mathbf{X}_t = (x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-d\tau})$  es un vector en el espacio de reconstrucción. Dichos algoritmos propuestos se pueden clasificar en locales y globales, dependiendo de la estrategia utilizada para aproximar  $F$ . Los métodos globales utilizan simultáneamente toda la información del registro histórico para determinar dicha función, mientras que los locales sólo emplean información contenida en el entorno próximo al estado actual del sistema para aproximar  $F$  en esa región. Como ejemplo de algoritmo global consideraremos a las Redes Neuronales (RN)<sup>4</sup>, las cuales realizan una exploración de todos los datos disponibles durante el proceso de aprendizaje (correspondiente al ajuste de sus parámetros internos minimizando una función error estipulada). Como algoritmo local utilizaremos una variante del

método llamado Aproximación de Hiperplano Local<sup>5,1</sup> (AHL), que básicamente consiste en buscar el hiperplano que mejor ajusta los valores que la función toma en la vecindad del punto a predecir. Para evaluar la performance de estos métodos usaremos el Error Cuadrático Medio Normalizado ( "Average Relative Variance")  $ARV = \sum_{t \in S} (x_t - \hat{x}_t)^2 / \sum_{t \in S} (x_t - \bar{x}_t)^2$ , donde  $\bar{x}_t$  es el promedio de los registros,  $\hat{x}_t$  es la predicción para  $x_t$  y  $S$  puede ser el conjunto de ajuste o el de predicción.

## II. SERIES CAÓTICAS ESTUDIADAS

### A. Mapa Logístico

En el caso del mapa logístico se generaron 20.000 iteraciones a partir de la ley  $x_{t+1} = 3.8x_t(1 - x_t)$  y se descartaron los primeros 10.000 valores. Se seleccionaron tres registros de  $N=20$ , 100 y 250 puntos respectivamente, los que se utilizaron para predecir los últimos 1.000 puntos generados. Los valores de  $N$  elegidos corresponden a 'escasos', 'suficientes' y 'muchos' datos, de acuerdo a la caracte-

rización realizada en<sup>6</sup> de las longitudes de los registros históricos utilizados para predecir esta serie.

### B. Ecuación de Mackey-Glass discretizada

Se generaron 25.000 iteraciones de la serie de acuerdo a la ley discreta  $x(t + \Delta t) = \frac{2-b\Delta t}{2+b\Delta t}x(t) + \frac{a\Delta t}{2+b\Delta t} \left( \frac{x(t+\Delta t-T)}{1+x^c(t+\Delta t-T)} + \frac{x(t-T)}{1+x^c(t-T)} \right)$ , con los parámetros  $a = 0.2$ ,  $b = 0.1$ ,  $c = 10$ ,  $T = 17$ ,  $\Delta t = 1$  y condiciones iniciales  $x(0) = x(1) = \dots = x(16) = 0.5$ . Se descartaron los primeros 10.000 valores y de los siguientes 9.000 se eligieron conjuntos de  $N=200$  (escasos), 500 (suficientes) y 1.500 (muchos) puntos, muestreando la serie cada 6 unidades de tiempo (correspondiente a un 'delay'  $\tau = 6$ , en el espacio de reconstrucción). Con las últimas 6.000 iteraciones se generó un intervalo de predicción de 1.000 puntos.

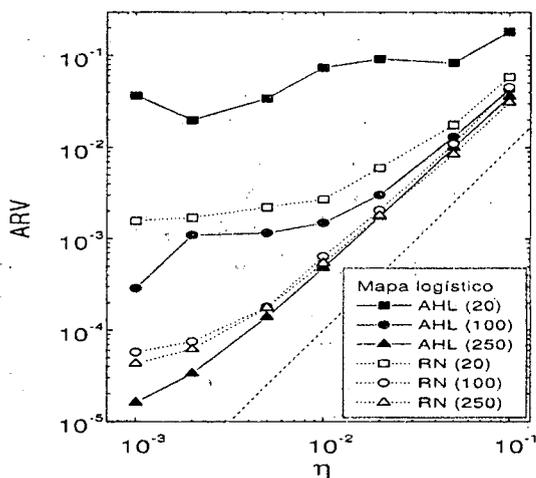


FIG. 1. Error de predicción para el Mapa Logístico en función de  $\eta$ .

### C. Series ruidosas

A los conjuntos generados se les adicionó ruido blanco  $\zeta_n$ , obteniéndose una nueva serie  $\{x_n\} = \{x_n + \lambda\zeta_n\}$ , donde el valor de  $\lambda$  se utiliza para graduar el nivel de ruido introducido. Para caracterizar este nivel se usó la relación señal-ruido  $\eta$ , definida en nuestro caso como la

relación entre los desvíos estándar de las dos series que se suman:

$$\eta = \frac{\sigma_r}{\sigma_s} \quad (1)$$

Aquí  $\sigma_r$  es el desvío estándar de la serie de ruido blanco y  $\sigma_s$  el de la serie original. Se utilizaron valores de  $\eta$  que van de 0.001 a 0.1.

La adición de ruido incorpora un límite máximo a la capacidad de predicción de cualquier método. Aún si el método fuese capaz de filtrar todo el ruido presente en los datos y de reproducir exactamente la dinámica original de la serie  $\{x_n\}$ , no podrá predecir el ruido adicionado a un valor de la serie ya que éste es elegido al azar. En esta situación ideal el ARV de predicción se reduciría a:

$$ARV = \frac{1}{\sigma_P^2 N} \sum_{k=1}^N (\hat{x}_k - x'_k)^2 = \frac{1}{\sigma_P^2 N} \sum_{k=1}^N (\lambda \zeta_n)^2 \quad (2)$$

Si además aproximamos el desvío estándar de la serie ruidosa por:

$$\sigma_P^2 \simeq \sigma_r^2 + \sigma_s^2 \quad (3)$$

e incorporamos (1), resulta:

$$ARV \simeq \frac{1}{(\sigma_r^2 + \sigma_s^2) N} \sum_{k=1}^N (\lambda \zeta_n)^2 = \frac{\eta^2}{1 + \eta^2} \quad (4)$$

expresión que da el límite inferior del error en función de  $\eta$ . Lo que analizaremos, por lo tanto, es cuánto se acerca el error de predicción de los distintos métodos a este nuevo límite.

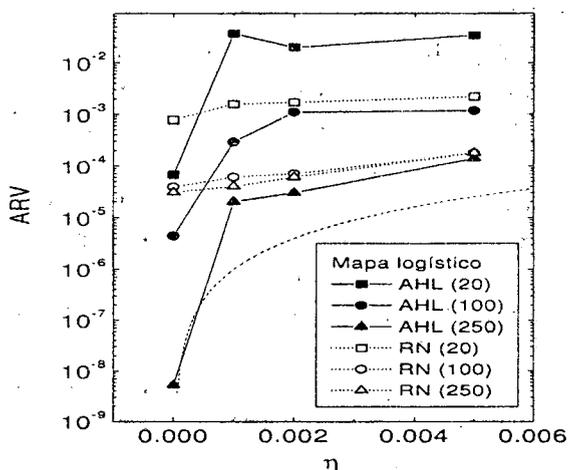


FIG. 2. Detalle de la Fig. 1 en escala lineal para valores pequeños de  $\eta$ .

### III. RESULTADOS

En la Fig. 1 se muestran los resultados generales para el mapa logístico. La Fig. 2 muestra un detalle de la figura anterior para valores de  $\eta < 0.01$ . El comportamiento de las RN en esta zona es prácticamente independiente del valor de  $\eta$ , lo que muestra la capacidad natural de las redes para filtrar el ruido presente. Las curvas que corresponden a 100 y 250 puntos son similares, un comportamiento similar al observado para series sin ruido<sup>1</sup>. La AHL, en cambio, muestra un rápido deterioro en su capacidad predictiva. Para 20 y 100 puntos las curvas se colocan rápidamente por encima de las correspondientes a las RN. Para 250 puntos, si bien el error se mantiene levemente por debajo del correspondiente a la red, para niveles muy bajos de ruido pierde la ventaja inicial de 4 órdenes de magnitud que presentaba en el caso de series puras ( $\eta = 0$ ). En la zona de mayor ruido ( $0.005 < \eta < 0.10$ ) las pendientes de todas las curvas son similares, dominadas por el aumento del límite mínimo de predicción (línea de trazos). En este caso, aún para 250 puntos las redes son competitivas con la AHL.

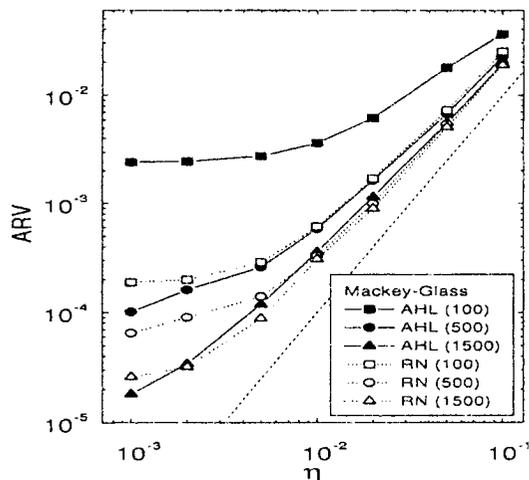


FIG. 3. Error de predicción para la ecuación de Mackey-Glass en función de  $\eta$ .

El mismo análisis se realizó para el mapa de Mackey-Glass. Los resultados pueden verse en la Fig. 3, que muestra un comportamiento general similar al observado en el caso anterior. Para este mapa, sin embargo, el deterioro de la performance del método local debido al

ruido es menos pronunciado que para el mapa logístico. En las situaciones de escasos y suficientes puntos, donde las redes eran superiores o equivalentes a la AHL, este comportamiento se mantiene para todos los niveles de ruido. Para el caso de muchos datos la AHL produce una mejor predicción que las RN aunque sólo para nulo o muy escaso ruido. En particular las curvas de error se cruzan para  $\eta = 0.002$ , mostrándose la red superior al método local para niveles mayores de ruido.

### IV. CONCLUSIONES

Hemos comparado la capacidad de dos buenos representantes de las metodologías locales y globales para la predicción de series temporales caóticas en presencia de ruido. No obstante su ventaja inicial en el caso de series puras y abundancia de datos<sup>1</sup>, el comportamiento de los métodos locales se deteriora rápidamente al introducir ruido, cosa que no ocurre con las RN. Éstas muestran una capacidad intrínseca de filtrado, que mantiene la calidad de su predicción para bajos niveles del mismo. El método local sólo se muestra superior al global en el caso de muchos puntos y escasísimo ruido presente en la señal.

<sup>1</sup> Verdes, P.F., Granitto, P.M., Navone, H.D. & Ceccatto, H.A., *Anales AFA* 1998, pág. 22.

<sup>2</sup> Farmer, J.D. & Sidorowich, J.J. *Phys. Rev. Lett.*, **59**, 845-848 (1987).

<sup>3</sup> Casdagli, M. & Eubank, S. eds. *Nonlinear Modeling and Forecasting*, Santa Fe Institute of Studies in the Sciences of Complexity Proc. vol XII (Redwood City: Addison-Wesley) (1992).

<sup>4</sup> Rumelhart, D.E., Hinton, G.E. & Williams, R.J. *Nature*, **323**, 533-536 (1986).

<sup>5</sup> Navone, H.D. & Ceccatto, H.A. *Journal of Physics A: Mathematical and General*, **28**, 3381-3388 (1995).

<sup>6</sup> Verdes, P.F., Granitto, P.M., Navone, H.D. & Ceccatto, H.A., en *Novel Intelligent Automation and Control Systems*, Vol. I, Ed: J. Pfeiffer, 129-145 (1998).