
Algoritmos de reducción de ruido de medida mediante técnicas de proyección local

M.E. Mera¹ y M. Morán²

Departamento de Análisis Económico I, Universidad Complutense.
Campus de Somosaguas, 28223 Madrid.

1 mera@ccee.ucm.es

2 mmoranca@ccee.ucm.es

El problema de reducción de ruido se plantea al registrar, usando un proceso de medición sujeto a error, una serie temporal generada por una dinámica determinista. Los algoritmos de reducción de ruido pretenden obtener una nueva serie que recupere en el mayor grado posible las propiedades geométricas, estadísticas o/y dinámicas de la serie temporal original.

A continuación revisamos los algoritmos de reducción de ruido de la literatura que parten de un enfoque no paramétrico, en los que la única hipótesis es la suavidad de la dinámica subyacente y que usan únicamente la información que proporciona la serie ruidosa. Todos estos algoritmos están basados en un análisis local de los datos, que será o bien una serie temporal escalar o bien una serie temporal vectorial.

Para el caso de series temporales vectoriales mostraremos que los algoritmos de reducción de ruido mejoran sustancialmente si se incorpora al esquema de reducción información acerca del grado de incertidumbre en cada una de las componentes de la serie temporal.

Finalmente veremos que gran parte de los algoritmos existentes pueden situarse en un marco común: se basan en proyecciones locales de los datos en subespacios lineales. En este marco, las diferencias entre los algoritmos se centran en cómo se eligen esos subespacios lineales, cómo se eligen los entornos y en la meta que los algoritmos persiguen.

1. Introducción

Muchos de los problemas de reducción de ruido pueden describirse por las ecuaciones

$$\begin{aligned}\mathbf{x}_{k+1} &= f(\mathbf{x}_k, \xi_k), \\ \mathbf{X}_k &= g(\mathbf{x}_k) + \mathbf{e}_k\end{aligned}$$

donde $\mathbf{x}_k \in \mathbb{R}^d$ es la variable de estado del sistema en el periodo k , f es una dinámica determinista suave, $g : \mathbb{R}^d \rightarrow \mathbb{R}^p$ es un observable suave, ξ_k, \mathbf{e}_k son variables aleatorias multidimensionales, y $\mathbf{X}_k, k = 1, 2, \dots, N$ es la serie temporal ruidosa observada. En este esquema $\mathbf{x}_k, f, g, \xi_k, \mathbf{e}_k$ son desconocidos, ξ_k se denomina ruido dinámico y \mathbf{e}_k ruido de medida.

Nosotros nos restringiremos al caso en el que no hay ruido dinámico, es decir

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k) \quad \text{y} \quad \mathbf{X}_k = g(\mathbf{x}_k) + \mathbf{e}_k,$$

y el ruido de medida \mathbf{e}_k es independiente e idénticamente distribuido.

Además adoptaremos un enfoque no paramétrico, no asumiremos ninguna forma específica para la dinámica subyacente f , solo se supondrá su suavidad, que es la hipótesis que permitirá aproximar f mediante ajustes lineales locales.

Hay muchos algoritmos en la literatura que parten de este enfoque (ver [1]-[8]). Pueden clasificarse en aquellos en los que el observable g es la función identidad y por tanto se tiene una serie temporal ruidosa vectorial, y aquellos diseñados para series temporales ruidosas escalares en los que $g : \mathbb{R}^d \rightarrow \mathbb{R}$ es un observable suave desconocido. Los algoritmos de este segundo tipo precisan de la reconstrucción de un espacio de estados multidimensional equivalente en algún sentido al espacio de estados original. El teorema de Takens garantiza [9] que para series sin ruido, para observables genéricos g , y para $m \geq 2d+1$, si se reconstruye en \mathbb{R}^m la serie escalar $\{x_k\}_{k=1, \dots, N} \in \mathbb{R}$ tomando las m -historias

$$\mathbf{x}_k^m := (x_k, x_{k+1}, \dots, x_{k+m-1})$$

el espacio reconstruido es difeomorfo al espacio de estados original y las principales propiedades de la dinámica f pueden inferirse de la dinámica shift $f^*(\mathbf{x}_k^m) := \mathbf{x}_{k+1}^m$.

La diferencia esencial entre las series ruidosas escalares y vectoriales es la estructura del término de error. En las series escalares se tiene que

$$X_k = g(\mathbf{x}_k) + e_k \in \mathbb{R},$$

el término de error e_k es unidimensional, luego al considerar la serie reconstruida

$$\begin{aligned} \mathbf{X}_k^m &:= (X_k, \dots, X_{k+m-1}) = (g(\mathbf{x}_k) + e_k, \dots, g(\mathbf{x}_{k+m-1}) + e_{k+m-1}) = \\ &(g(\mathbf{x}_k), \dots, g(\mathbf{x}_{k+m-1})) + (e_k, e_{k+1}, \dots, e_{k+m-1}) := \mathbf{x}_k^m + \mathbf{e}_k^m \end{aligned}$$

Por tanto el grado de incertidumbre de cada una de las m componentes del término de error \mathbf{e}_k^m para la serie reconstruida \mathbf{X}_k^m es el mismo, es decir \mathbf{e}_k^m es un error homocedástico (tiene matriz de varianzas-covarianzas $\Sigma = \sigma^2 I$ donde I es la matriz identidad $m \times m$).

Sin embargo en series vectoriales el término de error puede ser heterocedástico y pueden existir correlaciones entre sus componentes.

Los algoritmos diseñados para series escalares pueden utilizarse también para la reducción de ruido de series vectoriales. Sin embargo nosotros [10] hemos comprobado que los resultados mejoran sustancialmente si se incorpora al esquema de reducción de ruido información sobre las diferentes componentes del término de error.

Todos los algoritmos que describiremos a continuación están basados en proyecciones locales de los datos en subespacios adecuados.

Para el caso de series escalares la idea es que si no hubiera ruido la reconstrucción de la serie escalar original $\{\mathbf{x}_k^m, k = 1, \dots, N - m + 1\}$ se situaría en una variedad d -dimensional M de \mathbb{R}^m . Entonces para puntos \mathbf{x}_j^m en una bola U_i de radio pequeño centrada en \mathbf{x}_i^m , los vectores $\mathbf{x}_j^m - \langle \mathbf{x}_i^m \rangle_{U_i}$, donde $\langle \mathbf{x}_i^m \rangle_{U_i}$ es el centro de masas de U_i , generarían únicamente un subespacio lineal de d dimensiones, que es una aproximación al espacio tangente a M en $\langle \mathbf{x}_i^m \rangle_{U_i}$. Sin embargo el efecto del ruido hace que en la serie ruidosa reconstruida esos vectores diferencia $\mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i}$ generen todo \mathbb{R}^m . Una forma de eliminar parte del ruido que contienen los datos es proyectar esos vectores en subespacios lineales d -dimensionales adecuados (ver Figura 1).

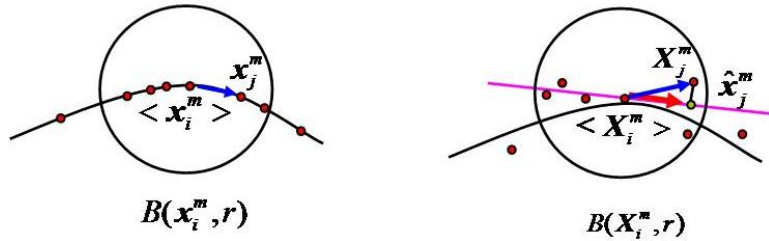


Figura 1. La serie limpia se sitúa en una variedad d -dimensional M de \mathbb{R}^m . La serie ruidosa, por efecto del ruido, genera todo \mathbb{R}^m . El ruido se elimina parcialmente proyectando los datos en cierto subespacio.

Para series temporales ruidosas vectoriales las dos series (la determinista y la ruidosa) están en \mathbb{R}^d , luego proyectar los vectores $\mathbf{X}_j - \langle \mathbf{X}_i \rangle_{U_i}$ en un subespacio de menor dimensión no reduce el ruido. Sin embargo la misma idea que hemos visto para el caso de series escalares reconstruidas puede aplicarse a series vectoriales si éstas se reconstruyen en espacios de mayor dimensión. Esto es lo que hacen los escasos algoritmos existentes diseñados para el caso particular de series vectoriales. Hegger y Schreiber [3] consideran la serie de \mathbb{R}^{3d} dada por $\mathbf{Z}_i = (\mathbf{X}_{i-1}, \mathbf{X}_i, \mathbf{X}_{i+1}), i = 2, \dots, N - 1$ y nosotros [10] consideramos la serie en \mathbb{R}^{2d} dada por $\mathbf{Z}_i = (\mathbf{X}_i, \mathbf{X}_{i+1}), i = 1, \dots, N - 1$. Puesto que sin ruido los datos se situarían nuevamente en una subvariedad d -dimensional de los respectivos espacios de alojamiento, los algoritmos eliminan

parte del ruido proyectando los datos, situados en entornos de los puntos de la serie, en subespacios lineales d -dimensionales adecuados.

Por lo tanto la primera característica que diferencia a los distintos algoritmos (ya sean para series escalares o vectoriales) es el espacio de alojamiento utilizado, que es el que determina los entornos que se van a considerar. La segunda característica es la métrica que se utilizará. Veremos que todos los algoritmos realizan de una u otra forma proyecciones en subespacios que son los óptimos con respecto de una métrica que será consistente con la meta que dicho algoritmo persigue, que es la tercera de las características distintivas de estos algoritmos.

La meta perseguida se refleja también en las medidas de reducción de ruido que los diferentes algoritmos considerarán. La reducción de ruido no se realiza en un único paso. La serie que el algoritmo proporciona en cada iteración será la serie ruidosa de entrada de la siguiente. Para decidir cuándo ya no es posible mejorar el resultado continuando con las iteraciones, se precisan medidas que cuantifiquen la cantidad de ruido que se ha eliminado. Esa medida de reducción estará relacionada con la meta que el algoritmo persigue.

2. Medidas de reducción de ruido

Aquí cabe distinguir las medidas de reducción de ruido que precisan del conocimiento de la serie determinista original y aquellas que no lo precisan. Es importante tener en cuenta que las medidas que precisan del conocimiento de la serie determinista solo usan esa información para medir la eficiencia de su algoritmo y no utilizan esa serie en el esquema de reducción de ruido. En las medidas que aparecen a continuación \mathbf{x} denota la serie limpia original, \mathbf{X} la serie ruidosa de partida y $\hat{\mathbf{x}}$ la serie que proporciona el algoritmo en cierta iteración.

Algunos autores [1]-[3], [6]-[8] pretenden reducir la distancia puntual

$$E(\mathbf{x}, \hat{\mathbf{x}}) := \left(\frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|^2 \right)^{1/2} \quad (1)$$

entre la serie determinista original \mathbf{x}_i y la serie ajustada $\hat{\mathbf{x}}_i, i = 1, \dots, N$, lo que es consistente con la meta de recuperar la serie original.

Otros autores [5] intentan maximizar la consistencia entre la serie ajustada y la dinámica lineal ajustada localmente, minimizando la cantidad

$$E_{dyn}(\hat{\mathbf{x}}) := \left(\frac{1}{N-1} \sum_{i=1}^{N-1} \|\hat{\mathbf{x}}_{i+1} - \hat{f}_i(\hat{\mathbf{x}}_i)\|^2 \right)^{1/2}, \quad (2)$$

donde \widehat{f}_i es una estimación lineal local para la dinámica f en \mathbf{x}_i . Esta idea es consistente con la idea de recuperar las propiedades dinámicas de la serie original. Obsérvese que esta medida de reducción no precisa el conocimiento de la serie determinista original. Cuando la dinámica f es conocida algunos autores consideran también la medida de reducción análoga que se obtiene sustituyendo en (2) las estimaciones de la dinámica por la propia dinámica, es decir consideran la cantidad

$$\left(\frac{1}{N-1} \sum_{i=1}^{N-1} \|\widehat{\mathbf{x}}_{i+1} - f(\widehat{\mathbf{x}}_i)\|^2 \right)^{1/2}.$$

Nosotros hemos propuesto [10] un enfoque alternativo. Pretendemos recuperar las propiedades geométricas y la regularidad estadística a largo plazo de la dinámica original. Para ello definimos la siguiente cantidad que es una medida de la distancia media de la serie ajustada al atractor

$$d_m(\widehat{\mathbf{x}}, \mathbf{x}) := \frac{1}{N} \sum_{j=1}^N \min_{i=1, \dots, N} \|\widehat{\mathbf{x}}_j - \mathbf{x}_i\| \quad (3)$$

La motivación de esta medida de la distancia entre dos series es la siguiente. Una órbita del sistema dinámico contiene (con probabilidad 1) toda la información acerca de la geometría, dinámica y comportamiento a largo plazo del sistema. Si consideramos dos órbitas del mismo sistema dinámico, la distancia puntual entre esas dos órbitas puede ser muy grande, y sin embargo con probabilidad 1 ambas contienen toda la información relevante acerca del comportamiento del sistema. Por ello se trata de definir una medida que asigne un valor arbitrariamente cercano (cuando la longitud de la órbita tiende a infinito) a esas dos órbitas. Una distancia de Hausdorff entre las dos series sí tiene esa propiedad. Para ello se considera

$$d(\widehat{\mathbf{x}}, \mathbf{x}) = \min_{i=1, \dots, N} \|\widehat{\mathbf{x}}_j - \mathbf{x}_i\|$$

Puesto que $d(\widehat{\mathbf{x}}, \mathbf{x}) \neq d(\mathbf{x}, \widehat{\mathbf{x}})$ se toma como distancia entre las dos series $d_H = \max\{d(\widehat{\mathbf{x}}, \mathbf{x}), d(\mathbf{x}, \widehat{\mathbf{x}})\}$. Puesto que la discrepancia en solo un punto de las dos series puede hacer que la distancia d_H sea grande, consideramos esta distancia conjuntamente con el promedio dado por (3).

El porcentaje de reducción asociado a cada una de estas medidas es $R_0 := 100(1 - \frac{E(\mathbf{x}, \widehat{\mathbf{x}})}{E(\mathbf{x}, \mathbf{X})})$, $R_{dyn} := 100(1 - \frac{E_{dyn}(\mathbf{x}, \widehat{\mathbf{x}})}{E_{dyn}(\mathbf{x}, \mathbf{X})})$ y $R_m := 100(1 - \frac{d_m(\widehat{\mathbf{x}}, \mathbf{x})}{d_m(\mathbf{X}, \mathbf{x})})$ respectivamente.

3. Adecuación de la estimación mínimo-cuadrática y métodos alternativos.

La mayoría de los algoritmos de reducción de ruido se basan en proyecciones locales de los datos o en aproximaciones lineales locales para la dinámica

subyacente obtenidas mediante el método de mínimos cuadrados. Puesto que la estimación mínimo cuadrática se obtiene proyectando el vector/matriz de variables dependientes en el subespacio generado por las columnas de la matriz cuyas entradas son los valores de las variables independientes, también es un método basado en una proyección local. Por ello las diferencias entre los diferentes métodos se centrarán en cómo se define ese mejor subespacio en el que proyectar y cuáles son los datos que se proyectan.

Veamos en un ejemplo las deficiencias del método de los mínimos cuadrados, que son las que motivan el considerar métodos alternativos. Consideremos una nube de puntos (x_i, y_i) , $i = 1, \dots, n$ en el plano generada por el modelo lineal $y = ax + b$ donde a y b son constantes desconocidas.

Si la variable dependiente y se registra con error, se tienen los valores (x_i, Y_i) , $i = 1, \dots, n$ donde $Y_i = y_i + e_{y_i}$ (ver la Figura 2(a)). Si estos errores son independientes e idénticamente distribuidos para todas las observaciones, tiene sentido el ajuste mínimo cuadrático

$$\min_{a,b} \frac{1}{n} \sum_{i=1}^n (Y_i - a - bx_i)^2$$

que proporciona estimadores insesgados y consistentes y que son los de máxima verosimilitud si la distribución del error es normal. Sin embargo si tanto las variables dependientes como las independientes están medidas con error entonces se tiene la nube (X_i, Y_i) con $X_i = x_i + e_{x_i}$, $Y_i = y_i + e_{y_i}$ y el ajuste mínimo cuadrático anterior proporciona estimados sesgados e inconsistentes. Si los términos de error tienen la misma varianza, el ajuste por mínimos cuadrados totales (también se denomina regresión ortogonal)

$$\min_{a,b,x_i} \frac{1}{n} \sum_{i=1}^n [(X_i - x_i)^2 + (Y_i - a - bx_i)^2]$$

proporciona estimadores insesgados, consistentes y de máxima verosimilitud si los errores son normales (ver Figura 2(b)). No es difícil ver que estos estimadores pueden obtenerse también resolviendo el problema

$$\min_T \frac{1}{n} \sum_{i=1}^n (Z_i - P_T Z_i)' (Z_i - P_T Z_i)$$

donde $Z_i = (X_i, Y_i)$ y $P_T Z_i$ denota la proyección ortogonal sobre la variedad afín unidimensional T y el mínimo se toma sobre el conjunto de todas las variedades afines unidimensionales.

Si las varianzas son distintas o existe correlación entre los errores entonces un estimador más adecuado se obtiene resolviendo

$$\min_T \frac{1}{n} \sum_{i=1}^n (Z_i - P_T Z_i)' \Sigma^{-1} (Z_i - P_T Z_i)$$

donde Σ es la matriz de varianzas covarianzas de los errores. En la figura 2 (c) la varianza del término de error en la variable x es mucho menor que la varianza en el término de error en la variable y y por tanto se proyecta con respecto de una métrica que da mucho más peso al valor observado de x que al valor observado de y . Esto también indica que el comportamiento del ajuste mínimo cuadrático clásico será mucho peor cuanto mayor sea la varianza del error en las variables independientes.

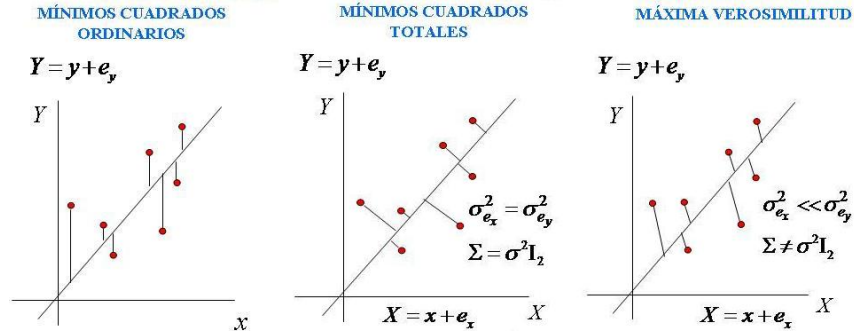


Figura 2. Nube de puntos ruidosa generada con un modelo lineal. En (a) se añade un ruido de observación a la variable dependiente y , en (b) se añade un ruido con la misma varianza a ambas variables, y en (c) la varianza del ruido añadido a la variable y es mucho mayor que la de la variable x .

En los problemas de reducción ruido, cuando se intenta aproximar localmente la dinámica subyacente, se tiene que tanto las variables dependientes como independientes están registradas con error, y por tanto el ajuste mínimo cuadrático no es adecuado. Si la serie es escalar, hemos visto que las diferentes componentes del error de la serie reconstruida son incorreladas y tienen la misma varianza, por lo que un ajuste mediante mínimos cuadrados totales sí sería adecuado, que es el que utilizan muchos de los algoritmos existentes. Sin embargo para series vectoriales las diferentes componentes del error pueden estar correladas y pueden tener diferente grado de incertidumbre, por lo que utilizar una métrica que incorpore esa información hará que los resultados del algoritmo mejoren sustancialmente (ver [10]).

3.1. Mejores subespacios lineales para unos datos

El siguiente teorema muestra cómo obtener el mejor subespacio lineal d -dimensional para unos datos en \mathbb{R}^n con respecto de una métrica dada. La demostración de este resultado puede verse en [10].

Sea A una matriz $n \times n$ simétrica y definida positiva, y sea \mathcal{L}_p el conjunto de todos los subespacios lineales p -dimensionales de \mathbb{R}^n . La proyección ortogonal

de $\mathbf{v} \in \mathbb{R}^n$ en $T \in \mathcal{L}_p$ con respecto de la métrica $\delta_{A^{-1}}$ está dada por

$$P_T \mathbf{v} := \arg \min_{\mathbf{w} \in T} \mathbf{v} A^{-1} \mathbf{w}^t.$$

Diremos que el subespacio lineal \mathcal{T}_p es el mejor subespacio en \mathcal{L}_p , para los puntos $\{\mathbf{Z}_k, k = 1, \dots, NV\} \in \mathbb{R}^n$, con respecto de la métrica $\delta_{A^{-1}}$, si minimiza en $T \in \mathcal{L}_p$ la cantidad

$$\mathcal{E}(T) := \frac{1}{NV} \sum_{i=1}^{NV} (\mathbf{Z}_k - P_T \mathbf{Z}_k)^t A^{-1} (\mathbf{Z}_k - P_T \mathbf{Z}_k).$$

Este mejor subespacio puede obtenerse a partir de los autovectores, en la métrica δ_A , de la matriz $n \times n$ definida por $M_{\mathbf{Z}\mathbf{Z}} := \frac{1}{NV} \sum_{k=1}^{NV} \mathbf{Z}_k \mathbf{Z}_k^t$.

Definimos a continuación qué es un conjunto ortonormal de autovectores de una matriz (que nuestro caso es $M_{\mathbf{Z}\mathbf{Z}}$) con respecto de una métrica dada.

Definition 1. *Un sistema de vectores $\{\mathbf{w}_1, \dots, \mathbf{w}_n\} \subset \mathbb{R}^n$ se dice que es un sistema ortonormal de autovectores para $M_{\mathbf{Z}\mathbf{Z}}$ en la métrica δ_A si satisfacen:*
(i) existen números reales λ_i (que se denominan autovalores para $M_{\mathbf{Z}\mathbf{Z}}$ en la métrica δ_A) tal que

$$M_{\mathbf{Z}\mathbf{Z}} \mathbf{w}_i = \lambda_i A \mathbf{w}_i, 1 \leq i \leq n,$$

(ii) $\mathbf{w}_i^t A \mathbf{w}_j = \delta_{ij}$, para todo i, j .

Teorema 1. *Sean $\lambda_1 \geq \dots \geq \lambda_n$ los autovalores de la matriz $M_{\mathbf{Z}\mathbf{Z}}$ en la métrica δ_A , $\{w_i, i = 1, \dots, n\}$ el correspondiente conjunto de ortonormal de autovectores y, para $1 \leq p \leq n$, sea B la matriz de dimensiones $n \times (n - p)$ cuyas columnas son los últimos $n - p$ autovectores $\{w_{p+1}, \dots, w_n\}$. Entonces, el mejor subespacio lineal en \mathcal{L}_p con respecto de la métrica $\delta_{A^{-1}}$ está dado por*

$$T_p := \text{span}\{Aw_1, \dots, Aw_p\}$$

y

$$P_{T_p} Z := (I - ABB^t)Z, \quad Z \in \mathbb{R}^n.$$

Además en Fuller (ver [11]) se prueba que si el modelo subyacente es lineal estos estimadores son insesgados y consistentes, y los de máxima verosimilitud si los errores son normales.

4. Algoritmos para series escalares

4.1. Método de Cawley-Hsu-Sauer [1],[6]

Toman como espacio de alojamiento \mathbb{R}^m para un valor de m adecuado, y consideran la métrica Euclídea para elegir los mejores subespacios lineales, es

decir toman la matriz identidad $m \times m$ como matriz A en el teorema anterior, lo que equivale a un ajuste mediante mínimos cuadrados totales.

Sea U_i una bola de radio pequeño centrada en \mathbf{X}_i^m y sea $\langle \mathbf{X}_i^m \rangle_{U_i}$ el centro de masas de los puntos pertenecientes a U_i . El algoritmo reduce el ruido proyectando los datos

$$\mathbf{Z}_j := \frac{\mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i}}{\left\| \mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i} \right\|}, \quad \text{para } \mathbf{X}_j^m \in U_i$$

en el mejor subespacio lineal d -dimensional \mathcal{T}_d . Así una de las diferencias de este algoritmo con respecto de los restantes de la literatura es ponderar de igual manera los puntos cercanos y los lejanos al centro de masas (al considerar vectores normalizados). Los autores lo justifican diciendo que si se consideraran los vectores sin normalizar se estaría dando mayor peso a los vectores de mayor norma (es decir a los puntos lejanos) lo que, debido a la curvatura y al ruido, introduciría un sesgo no deseado en la estimación de ese mejor subespacio (que pretende ser una estimación al espacio tangente).

Finalmente en vez de considerar simplemente la proyección de los datos $\mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i}$ en ese mejor subespacio \mathcal{T}_d , dan como estimación de \mathbf{x}_j el valor

$$\hat{\mathbf{x}}_j = \langle \mathbf{X}_i^m \rangle_{U_i} + \alpha(\mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i}) + (1 - \alpha)P_{\mathcal{T}_d}(\mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i})$$

donde $0 \leq \alpha < 1$.

La inclusión del término $\alpha(\mathbf{X}_j^m - \langle \mathbf{X}_i^m \rangle_{U_i})$ en la expresión anterior impide que la serie ajustada y la original puedan separarse mucho. Esto está en concordancia con la meta de la reducción puntual. Los autores justifican la inclusión de este término con el fin suavizar las correcciones excesivamente grandes debidas a los efectos de la curvatura.

Otra de las características distintivas de este algoritmo es que los entornos se eligen disjuntos, lo que evita el problema de tener diferentes estimaciones para el mismo punto de la serie reconstruida. Finalmente se toma como estimación de la serie escalar aquella para la que su reconstrucción m -dimensional está más próxima a la serie vectorial ajustada por el algoritmo. Cada elemento de la serie escalar, a excepción de los $m - 1$ primeros y $m - 1$ últimos valores de la serie, aparece como componente en m -vectores consecutivos de la serie reconstruida, y esa mejor serie escalar en el sentido mínimo cuadrático se obtiene tomando el promedio de esos m valores. Esta mejor serie escalar será la serie de entrada en la siguiente iteración del algoritmo. La medida de reducción de ruido que utilizan para decidir cuándo debe parar el algoritmo está basada en la distancia puntual entre la serie original y la serie ajustada. Para los casos donde no se dispone de la serie determinista original proponen una medida del grado de reducción independiente de la serie determinista y que se basa en las correcciones que realiza el algoritmo iteración a iteración.

4.2. Método de Schreiber-Grassberger [8]

El algoritmo primero aloja la serie escalar en \mathbb{R}^m para un valor adecuado (impar) de m . Luego asigna como estimación de la coordenada central de cada vector m -dimensional \mathbf{X}_i^m una función lineal de las restantes $m - 1$ componentes del vector. Esta función lineal se estima usando una bola U_i centrada en \mathbf{X}_i^m y de radio pequeño. Esto proporciona una única estimación para cada uno de los puntos de la serie escalar, con lo que se evita el problema de las correcciones múltiples y cómo obtener a partir de ellas una única estimación para el correspondiente valor de la serie escalar.

En este algoritmo la coordenada central de cada m -historia juega un papel esencial, los autores justifican la elección de esta coordenada de la siguiente manera: si la coordenada estimada a partir de las restantes fuera la última, a medida que se van haciendo las correcciones y usando éstas para realizar nuevas correcciones, la naturaleza caótica de la dinámica haría que esas correcciones fuesen cada vez mayores, al menos para las direcciones correspondientes a la variedad inestable. Si la estimación fuera de la primera coordenada a partir de las restantes, por la misma razón las correcciones estarían mal condicionadas en la variedad estable. La manera de minimizar este problema es incorporar información acerca del pasado y del futuro, lo que se consigue si es la coordenada central la que se estima linealmente a partir de las restantes.

La función lineal que proporciona la coordenada central a partir de las restantes coordenadas, puede interpretarse en nuestro esquema como una proyección de \mathbf{X}_i^m en el mejor subespacio lineal $m - 1$ dimensional con respecto de una métrica $\delta_{A^{-1}}$ que correspondería a una matriz diagonal A con todas sus entradas nulas a excepción de la central que sería igual a uno. Puesto que esa matriz no sería invertible bastaría considerarla como límite de una sucesión de matrices con entradas en la diagonal tendiendo a cero a excepción de la entrada central de la diagonal que sería siempre uno (en los resultados numéricos proporcionados en este artículo toman el valor 0,001 para estas entradas y uno en el elemento central de la diagonal). Como medidas de reducción de ruido consideran tanto la distancia puntual entre la serie original y la ajustada como el grado de adecuación de la serie ajustada a las funciones lineales estimadas por el algoritmo dada por la expresión (2).

Merece la pena destacar el sencillo pero eficiente algoritmo de reducción de ruido propuesto por Schreiber [7] que procede de la misma manera descrita anteriormente, pero tomando como estimación de la coordenada central la media de las coordenadas centrales para los puntos del entorno. Este es el caso donde en vez de tomar una función lineal de las restantes coordenadas se toma una función constante.

4.3. Método de Grassberger et al. [2]

Este algoritmo es una modificación del anterior, que en vez de modificar la coordenada central de cada m -historia corrige varias de las coordenadas

centrales. Cuántas se corrigen y la dimensión de los subespacios en los que se proyecta son parámetros del algoritmo. Puesto que cada dato de la serie escalar es alguna de esas coordenadas centrales de varias m -historias, la estimación final es un promedio de dichos valores. En nuestro marco, la métrica $\delta_{A^{-1}}$ es la dada por una matriz diagonal A con todas sus entradas iguales a cero a excepción de esos valores centrales que toman el valor uno. Nuevamente esta matriz puede considerarse el límite de una sucesión de matrices diagonales con valores centrales iguales a uno y valores extremos tendiendo a cero. En los experimentos numéricos que proporcionan estos autores se toman los dos valores extremos de la diagonal iguales a 0,001 y los centrales iguales a uno. Como medidas de reducción de ruido usan tanto R_0 como R_{dyn} .

En [2] se comparan los resultados, que para las dinámicas de Henon, Lorenz e Ikeda, y para diferentes niveles de ruido, consiguen los diferentes algoritmos diseñados para series escalares. Los resultados indican la superioridad del algoritmo propuesto por estos autores.

5. Algoritmos para series vectoriales.

5.1. Método de Kostelich-Yorke[5]

La diferenciabilidad de $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ implica que, para cualquier punto \mathbf{x} perteneciente a una bola de radio pequeño centrada en un punto \mathbf{x}_i del atractor, se cumple que

$$f(\mathbf{x}) \sim \mathbf{B}_i \mathbf{x} + \mathbf{d}_i,$$

donde \mathbf{B}_i es una matriz $d \times d$, que es una estimación de la matriz jacobiana de f en \mathbf{x}_i , y \mathbf{d}_i es un vector d -dimensional. Por ello el algoritmo considera bolas U_i de radio pequeño, centradas en cada uno de los puntos \mathbf{X}_i de la serie temporal d -dimensional ruidosa y toma como estimación de \mathbf{B}_i y \mathbf{d}_i la solución del problema de estimación mínimo cuadrático

$$\min_{\mathbf{B}, \mathbf{d}} \sum_{j \in U_i} \|\mathbf{X}_{j+1} - (\mathbf{B}\mathbf{X}_j + \mathbf{d})\|^2$$

Una primera estimación $\hat{\mathbf{x}}_i$ para \mathbf{x}_i sería $\hat{\mathbf{x}}_i := \mathbf{B}_i \mathbf{X}_i + \mathbf{d}_i$, pero esta estimación está demasiado influenciada por el ruido que incorpora \mathbf{X}_i así como por los errores en la estimación de \mathbf{B}_i y \mathbf{d}_i . Por esta razón el algoritmo ajusta una nueva serie temporal que es consistente con los ajustes lineales realizados anteriormente. Esto se realiza seleccionando ventanas de p puntos consecutivos $\{\mathbf{X}_i, \mathbf{X}_{i+1}, \dots, \mathbf{X}_{i+p}\}$ para tomar como estimación de éstos los valores $\{\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i+1}, \dots, \hat{\mathbf{x}}_{i+p}\}$ que resuelven el problema de optimización

$$\min_{\hat{\mathbf{x}}_i, \dots, \hat{\mathbf{x}}_{i+p}} w \sum_{j=i}^{i+p} \|\hat{\mathbf{x}}_j - \mathbf{X}_j\|^2 + \sum_{j=i}^{p-1} \|\hat{\mathbf{x}}_{j+1} - (\mathbf{B}_j \hat{\mathbf{x}}_j + \mathbf{d}_j)\|^2 \quad (4)$$

donde w es un factor de ponderación, que fuerza a que la serie ajustada no se aleje demasiado de la serie de entrada del algoritmo. Así el método resultante es consistente con la obtención de una serie que emule en lo posible la dinámica de la serie original (ver (2)) y con reducir la distancia puntual entre la serie ajustada y la serie determinista original.

Este algoritmo fue propuesto inicialmente para series escalares. En este caso el último de los pasos del algoritmo se adapta para que la serie resultante sea una serie escalar.

La conexión entre este método y los métodos basados en proyecciones viene del hecho de que el ajuste mínimo cuadrático es una proyección sobre un subespacio lineal adecuado (el generado por las columnas linealmente independientes de la matriz de observaciones correspondiente a las variables independientes).

5.2. Método de Hegger-Schreiber[3]

Es una adaptación para series vectoriales del algoritmo, para series escalares, de Schreiber y Grassberger [8] descrito en la sección anterior. Con el fin de incorporar la información acerca del pasado y del futuro, y que por tanto la estimación esté bien condicionada tanto en la variedad estable como en la inestable, el algoritmo considera la serie reconstruida

$$\mathbf{Z}_i := (\mathbf{X}_{i-1}, \mathbf{X}_i, \mathbf{X}_{i+1}) \in \mathbb{R}^{3d}, i = 2, \dots, N - 1.$$

Así el espacio de trabajo es \mathbb{R}^{3d} , donde d es la dimensión del espacio original.

La hipótesis es que la serie temporal limpia satisface, para puntos \mathbf{z}_j cercanos a $\mathbf{z}_i := (\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})$, la relación lineal

$$\mathbf{x}_j \sim \mathbf{B}_i \mathbf{x}_{j-1} + \mathbf{C}_i \mathbf{x}_{j+1} + \mathbf{d}_i$$

donde \mathbf{B}_i y \mathbf{C}_i son matrices $d \times d$ y \mathbf{d}_i es un vector d -dimensional. Este ajuste lineal se obtiene resolviendo el problema de optimización mínimo cuadrático

$$\min_{\mathbf{B}, \mathbf{C}, \mathbf{d}} \sum_{j: \mathbf{x}_j \in U_i} \|\mathbf{X}_j - (\mathbf{B}\mathbf{X}_{j-1} + \mathbf{C}\mathbf{X}_{j+1} + \mathbf{d})\|^2$$

donde U_i es una bola de radio pequeño centrada en \mathbf{Z}_i . Sea

$$\mathbf{X}_i^{corr} := \mathbf{B}_i \mathbf{X}_{i-1} + \mathbf{C}_i \mathbf{X}_{i+1} + \mathbf{d}_i$$

la estimación de \mathbf{x}_i que proporciona el algoritmo en este primer paso. Entonces la estimación final que proporciona el algoritmo para \mathbf{x}_i es

$$\hat{\mathbf{x}}_i = (1 - \alpha)\mathbf{X}_i + \alpha\mathbf{X}_i^{corr}$$

donde $0 < \alpha \leq 1$, con lo que se consigue que la serie ajustada no se aleje demasiado de la serie de entrada.

Puede probarse que \mathbf{X}_i^{corr} son las coordenadas centrales de la proyección ortogonal de \mathbf{Z}_i en el mejor subespacio lineal d -dimensional con respecto de una métrica $\delta_{A^{-1}}$ donde A es el límite de una sucesión de matrices diagonales que tienen unos en las d coordenadas centrales de la diagonal y valores tendiendo a cero en los $2d$ valores de los extremos.

Los autores miden el grado de reducción en términos de la distancia puntual entre la serie ajustada y la original no ruidosa, y en términos de adecuación de la serie ajustada a la dinámica lineal ajustada.

5.3. Método de Mera-Morán[10]

El algoritmo que nosotros proponemos persigue una reducción de ruido en el sentido estadístico, y es el único algoritmo en la literatura que incorpora en el esquema de reducción la información que contiene un ruido heterocedástico y correlado.

Su punto de partida es el algoritmo de Kostelich y Yorke [5]. Estos autores fueron los primeros en señalar la falta de adecuación del ajuste mínimo cuadrático en el problema de reducción de ruido, aunque no proporcionaron un método para resolver este problema. Se sabe que en modelos lineales donde tanto las variables dependientes como las independientes están sujetas a error, el ajuste mínimo cuadrático es inconsistente y sesgado, aumentando el sesgo a medida que aumenta la varianza del error en las variables independientes. Fuller [11] proporcionó la teoría adecuada para este tipo de modelos (modelos lineales con errores de medida en las variables independientes) mostrando cómo obtener estimadores insesgados, consistentes y que, bajo hipótesis de normalidad en la distribución de los errores, son los de máxima verosimilitud.

Nosotros hemos incorporado esta teoría en nuestro algoritmo y hemos demostrado que consiste en un método de proyección con respecto de una métrica adecuada.

Nuestro algoritmo usa como espacio de trabajo \mathbb{R}^d donde d es la dimensión del espacio en el que está definida la dinámica original. Para cada punto \mathbf{X}_i de la serie se toma una bola U_i de radio pequeño centrada en dicho punto, y utilizando la teoría desarrollada por Fuller, se obtiene un estimador consistente e insesgado para la aplicación lineal cuya matriz \mathbf{B}_i satisface

$$\mathbf{X}_{j+1} - \langle \mathbf{X}_{i+1} \rangle_{U_i} \approx \mathbf{B}_i (\mathbf{X}_j - \langle \mathbf{X}_i \rangle_{U_i}), \mathbf{X}_j \in U_i$$

donde $\langle \mathbf{X}_i \rangle_{U_i}$ denota el centro de masas de los puntos de la bola U_i y $\langle \mathbf{X}_{i+1} \rangle_{U_i}$ es el centro de masas de sus imágenes. Además esta teoría proporciona, para cada $\mathbf{X}_j \in U_i$, la estimación máximo verosímil $\hat{\mathbf{z}}_j$ de

$$\mathbf{z}_j := (\mathbf{x}_{j+1} - \langle \mathbf{x}_{i+1} \rangle_{U_i}, \mathbf{x}_j - \langle \mathbf{x}_i \rangle_{U_i}).$$

El haber sido capaces de obtener esta estimación hace innecesario el último paso (ver 4) del algoritmo de Kostelich y Yorke en el que se buscaba una serie consistente con la dinámica.

La conexión entre nuestro algoritmo y los basados en proyecciones viene del hecho de que $\hat{\mathbf{z}}_j$ coincide (ver [10]) con la proyección ortogonal de \mathbf{Z}_j en el mejor subespacio lineal d -dimensional en la métrica Σ^{-1} , donde Σ es la matriz de covarianzas de los errores \mathbf{e}_j en $\mathbf{Z}_j := \mathbf{z}_j + \mathbf{e}_j$. Así nuestro método utiliza la información que la matriz de covarianzas de los errores contiene acerca del error para reducir el ruido.

Como en general la matriz de covarianzas Σ de los errores es desconocida, se toma en una primera iteración del algoritmo como estimación de ésta la matriz identidad, estimación que es corregida en las siguientes iteraciones considerando la matriz de covarianzas empírica de las estimaciones del error que realiza el algoritmo.

La figura 3(a) muestra una serie temporal de 10000 puntos generada por la dinámica de Henon y la serie ruidosa (b) que se obtiene al añadir a ésta un ruido normal incorrelado y homocedástico con una desviación típica del 1% de la desviación típica de la señal ($\sigma_{\varepsilon_1} = \sigma_{\varepsilon_2} = 1\% \sigma_s$, donde σ_s es la desviación típica de la señal) y en (c) se representan juntas las series limpia y limpiada, donde puede verse que ambas series son prácticamente indistinguibles. La reducción de ruido en términos de distancia puntual es solo $R_0 = 40\%$, sin embargo en términos de distancia media de hausdorff es $R_m = 84\%$.

En la figura 4 se presenta una serie ruidosa de la misma dinámica pero con un ruido altamente heterocedástico: la desviación típica en el término de error que se añade a la primera de las componentes de la serie es $\sigma_{\varepsilon_1} = 1\% \sigma_s \approx 1\% \sigma_x$, y la desviación típica del término de error que se le añade a la segunda de las componentes de la serie es $\sigma_{\varepsilon_2} = 15\% \sigma_s \approx 50\% \sigma_x$. En la misma figura aparece la serie que proporciona nuestro algoritmo. Los niveles de reducción son $R_0 = 83\%$, $R_{dyn} = 97\%$ y $R_m = 94\%$ respectivamente.

Finalmente en la figura 5 pueden verse los resultados que se obtienen para una serie de solo 500 puntos generada de la misma forma que la de la figura 4 usando mínimos cuadrados totales y nuestro método, donde puede comprobarse la mejora sustancial que se obtiene al incorporar la información sobre las varianzas y correlaciones empíricas de cada una de las componentes del error, incluso en series de longitud muy corta.

6. Conclusiones

Hemos revisado varios algoritmos de reducción de ruido que adoptan un enfoque no paramétrico y basados en ajustes lineales. Todos estos algoritmos producen una nueva serie que se obtiene a través de algún proceso de proyección de la serie ruidosa original en algún subespacio adecuado. Los algoritmos difieren en los espacios ambiente en los que se trabaja, en la métrica elegida y en las metas que persiguen. Aunque éstas son las tres características distintivas más importantes, hay otras diferencias que se basan en los dispositivos específicos que cada uno de los algoritmos incorpora con el fin de mejorar

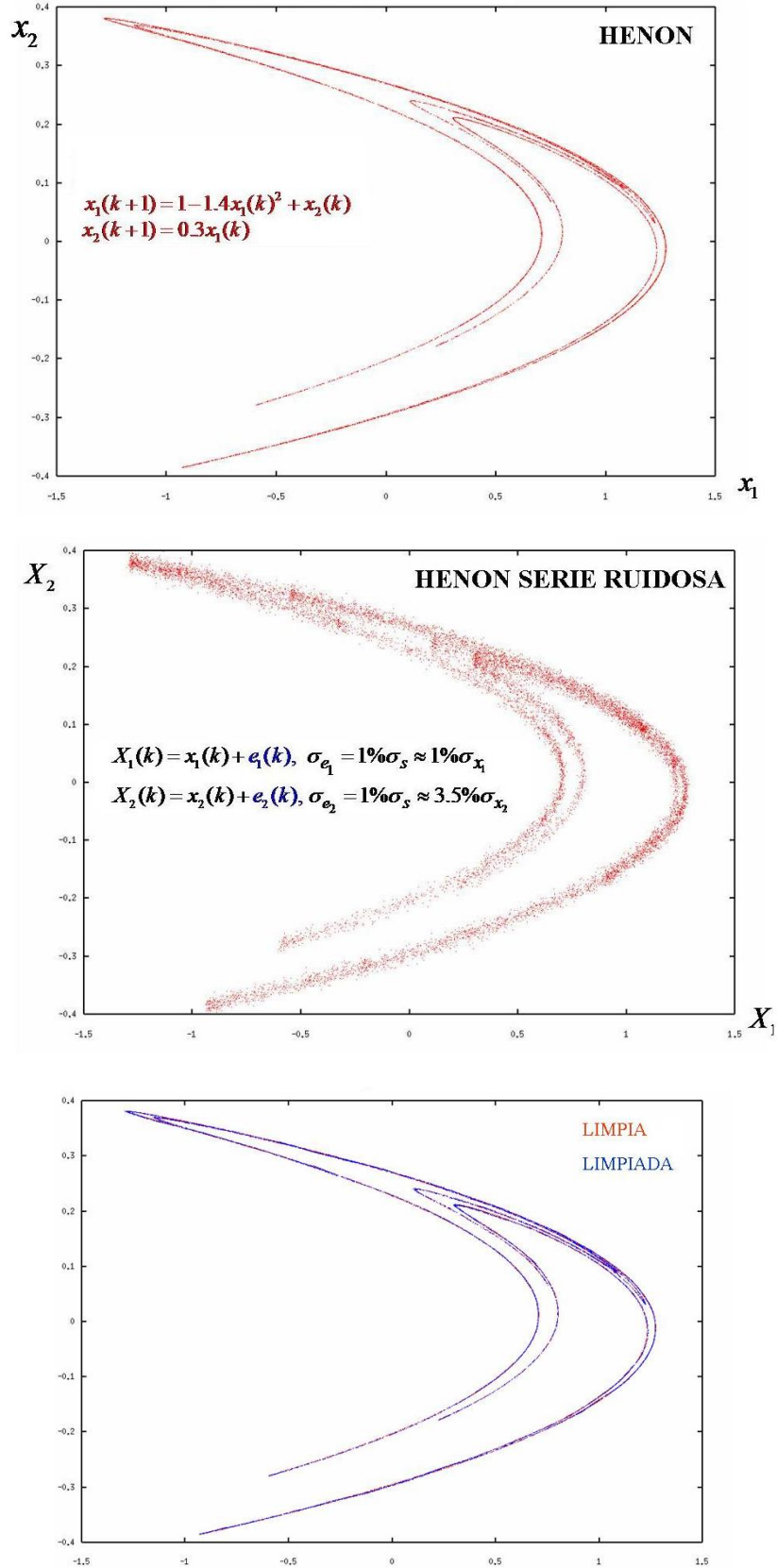


Figura 3. (a) Serie de 10000 puntos generada por la dinámica de Henon, (b) serie ruidosa con $\sigma_{\varepsilon_1} = \sigma_{\varepsilon_2} = 1\% \sigma_s$, y (c) serie original (en rojo) y serie ajustada por el algoritmo (en azul) en 16 iteraciones.

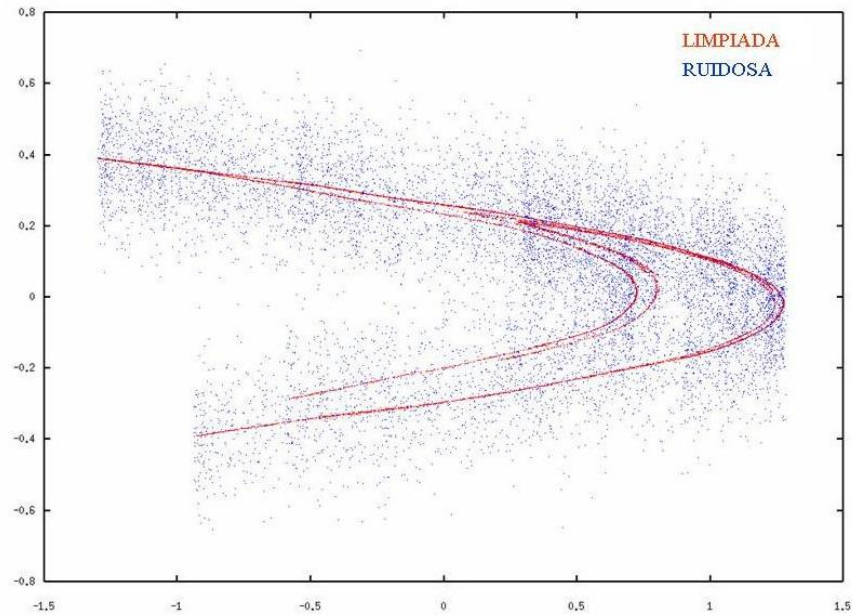


Figura 4. Serie ruidosa altamente heterocedástica (en azul) y serie ajustada por el algoritmo (en rojo) después de 6 iteraciones.

los resultados. Hay dispositivos para evitar correcciones excesivamente importantes, para seleccionar mejor el tamaño o la forma de los entornos [13], para que la dimensión del subespacio en el que se proyecta pueda variar de punto a punto [12], para que los efectos no lineales de la dinámica subyacente reduzcan su importancia y otras muchas.

Todos los algoritmos propuestos en la literatura presentan un buen comportamiento en relación con las metas que persiguen. Por ejemplo, para una serie temporal escalar de 5000 puntos de un Hénon al que se le ha añadido un ruido incorrelado y homocedástico de amplitud que varía entre un 1% y un 10% de la desviación típica de la señal, los diferentes algoritmos obtienen un porcentaje de reducción de ruido en términos de distancia puntual entre el 65% y el 73% y entre 83% y 87% en términos de autoconsistencia con la dinámica. Los resultados parecen indicar que la obtención de una serie temporal que converja a la original en términos de distancia puntual es poco factible salvo en los casos de ruidos de amplitud muy pequeña, y por ello la meta de la predictibilidad a corto plazo de series con ruidos es poco viable. Es mucho más realista la meta de la predictibilidad a largo plazo, donde incluso para series ruidosas con amplitud media parece factible la obtención de series con

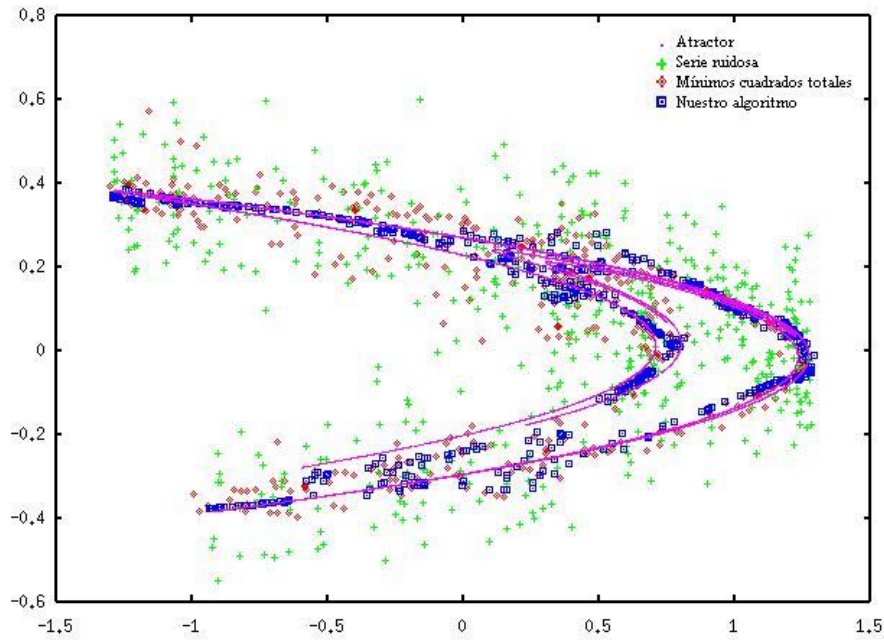


Figura 5. Serie de 500 puntos de un Henon con un ruido altamente heterocedástico (en verde) y resultados que se obtienen usando mínimos cuadrados totales (en rojo) y nuestro algoritmo (en azul).

propiedades geométricas, dinámicas y estadísticas muy similares a las de la serie original.

El caso de series ruidosas vectoriales ha sido escasamente tratado en la literatura, pues se había considerado que estas series podían analizarse con los algoritmos diseñados para series ruidosas escalares. Nuestros resultados demuestran que incorporar la información que proporciona la matriz de covarianzas de los errores en la métrica utilizada en el esquema de reducción, mejora sustancialmente la eficiencia del algoritmo.

Agradecimientos

Nos gustaría agradecer a los organizadores todo su esfuerzo y dedicación en la organización de estas jornadas, esperando que una iniciativa como esta pueda tener continuidad en el futuro.

Esta investigación ha sido financiada por el Ministerio de Educación y Ciencia a través del proyecto de investigación MTM2006-02372.

Referencias

1. R. Cawley and G-H. Hsu. Phys. Rev. A 46, 6, 3057-3082 (1992).
2. P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath and T. Schreiber. Chaos 3, 2, 127-141 (1993)
3. R. Hegger and T. Schreiber. Phys. Lett. A 170, 305-310 (1992).
4. E. J. Kostelich and T. Schreiber. Phys. Rev. E 48, 3, 1752-1763 (1993).
5. E. J. Kostelich and J. A. Yorke. Physica D 41, 183-196 (1990).
6. T. Sauer. Physica D 58, 193-201 (1992).
7. T. Schreiber. Phys. Rev. E 47, 4, 2401-2404 (1993)
8. T. Schreiber and P. Grassberger. Phys. Lett. A 160, 411-418 (1991).
9. F. Takens. "Dynamical Systems and turbulence." D. A. Rang and L. S. Young eds. Lecture Notes in Mathematics 898, 396 (1981).
10. M.E. Mera and M. Morán. Chaos 16, 1, 13116 (2006).
11. W. A. Fuller. "Measurement error models." John Wiley & Sons (1987).
12. A. Leontitsis, T. Bountis and J. Pagge. Chaos 14, 1, 106-11, (2004).
13. A. Kern, W.-H. Steeb and R. Stoop. Int. J. Mod. Phys. C 11, 1, 125-146 (2000).