




## Entrenamiento de redes neuronales recurrentes usando la dimensión de encaje en el espacio de las fases

*Training of recurrent neural networks using the embedding dimension in  
the space the phases*

- <sup>1</sup> Nelly Patricia Perugachi Cahueñas  <https://orcid.org/0000-0001-6331-9551>  
Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Administración de  
Empresas, Escuela de Transporte  
[patricia.perugachi@epoch.edu.ec](mailto:patricia.perugachi@epoch.edu.ec)
- <sup>2</sup> Mariela Micaela Moreno Palacios  <https://orcid.org/0000-0002-1605-6442>  
Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Ciencias, GEAA  
[mariela.moreno@epoch.edu.ec](mailto:mariela.moreno@epoch.edu.ec)
- <sup>3</sup> Arquímedes Xavier Haro Velasteguí  <https://orcid.org/0000-0003-3391-5082>  
Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Ciencias, Escuela de  
Física  
[aharo@epoch.edu.ec](mailto:aharo@epoch.edu.ec)



### Artículo de Investigación Científica y Tecnológica

Enviado: 17/05/2022

Revisado: 28/06/2022

Aceptado: 01/07/2022

Publicado: 29/07/2022

DOI: <https://doi.org/10.33262/concienciadigital.v5i3.1.2252>

**Cítese:** Perugachi Cahueñas, N. P., Moreno Palacios, M. M., & Haro Velasteguí, A. X. (2022).  
Entrenamiento de redes neuronales recurrentes usando la dimensión de encaje en el  
espacio de las fases. *ConcienciaDigital*, 5(3.1), 254-274.  
<https://doi.org/10.33262/concienciadigital.v5i3.1.2252>



**CONCIENCIA DIGITAL**, es una revista multidisciplinar, **trimestral**, que se publicará en soporte electrónico  
tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica  
que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva  
mediante su intervención cambios positivos en la sociedad. <https://concienciadigital.org>  
La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara  
Ecuatoriana de Libro con No de Afiliación 663) [www.celibro.org.ec](http://www.celibro.org.ec)

Esta revista está protegida bajo una licencia Creative Commons Attribution Non Commercial No Derivatives  
4.0 International. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

**Palabras claves:**

Teoría del Caos, redes neuronales recurrentes, dimensión de encaje, sistemas caóticos.

**Keywords:**

chaos theory, recurrent neural networks, embedding dimension, chaotic system.

**Resumen**

Se propone un método usando Redes Neuronales Recurrentes para la predicción de datos caóticos, aplicando la Teoría del Caos, para estudiar el comportamiento dinámico de los datos en el espacio multidimensional de las fases, establecer la correlación de los mismos y determinar la dimensión de encaje como base para el entrenamiento de las redes neuronales, así como determinar las características dinámicas del sistema calculando los coeficientes de Lyapunov y la entropía de Kolmorov-Siani, que nos indican el grado de desorden que tiene el sistema, para proyectar la precisión de la predicción. Se usan datos de contaminantes PM2.5 tomados en el centro Histórico de la ciudad de Quito, en intervalos de una hora, entre los años 2005 a 2019. Los resultados determinan que las series de datos corresponden a un sistema caótico (más de un coeficiente positivo de Lyapunov), por lo que se justifica la aplicación de la Teoría del Caos en el análisis de los mismos, dando buenos resultados en las predicciones aplicando los métodos de redes neuronales recurrentes de Elman y Jordan, al comparar las series predichas se demuestran que no presentan diferencias significativas entre ellas, ni con los datos medidos, usando el método de varianza con 0,05 de significancia, el error cuadrático porcentual respecto al rango de variación de los datos es aproximadamente del 5 % en ambos casos. **Objetivos:** Proponer un método que ayude al entrenamiento de las redes neuronales usando la Teoría del caos, mediante la implementación de la dimensión de encaje en el espacio de las fases.

**Abstract**

A method is proposed using Recurrent Neural Networks for the prediction of chaotic data, applying Chaos Theory, to study the dynamic behavior of the data in the multidimensional space of the phases, establish the correlation of the same and determine the embedding dimensions as a basis for the training of the neural networks, as well as determine the dynamic characteristics of the system by calculating the Lyapunov coefficients and the Kolmorov-Siani entropy, that tell us the degree of disorder that the system has, to project the accuracy of the prediction. Data on PM2.5 pollutants taken in the Historic Center of the city of Quito, at one-hour intervals, between the years 2005 to 2019, are used. The results determine that the data series correspond to a chaotic system (more

---

than one positive Lyapunov coefficient), so the application of Chaos Theory in the analysis of them is justified, giving good results in the predictions applying the methods of recurrent neural networks of Elman and Jordan, when comparing the predicted series they are shown that they do not present significant differences between them, nor with the measured data, using the method of variance with 0.05 significance, the percentage square error with respect to the range of variation of the data is approximately 5% in both cases. **Objectives:** To propose a method that helps the training of neural networks using Chaos Theory, by implementing the socket dimension in the space of the phases.

---

## Introducción

Uno de los problemas de las redes neuronales son los valores de entrenamiento de las series, lo cual se los puede tomar al azar, sin embargo, si se cambia el resultado puede ser diferente, por lo que en este trabajo se propone juntar la Teoría del Caos y las redes neuronales para determinar ese valor, tomando datos de PM2.5 de la Red de monitoreo Atmosférico de Quito (REMMAQ). En el año 2002 se implementó la Red Metropolitana de Monitoreo Atmosférico de Quito, para controlar los efectos de la contaminación del aire (Rodríguez, 2015).

La teoría del caos es una rama de las ciencias que se usa para estudiar sistemas complejos, sensibles a las condiciones iniciales, en los cuales pequeñas variaciones en estas condiciones implican grandes diferencias en el comportamiento futuro, imposibilitando la predicción a largo plazo. Esto sucede a pesar de que los sistemas son en rigor determinísticos, es decir; su comportamiento puede ser completamente determinado conociendo sus condiciones iniciales (Fowler & McGuinness, 2010).

La red neuronal tiene un proceso concreto y tiende a almacenar procedimientos prácticos y hacerlos valederos para su empleo (Rebala et al., 2019). Es comparable al cerebro del ser humano en dos perspectivas, la primera, la red obtiene conocimientos mediante un procedimiento de preparación; la segunda, los enlaces inter-neuronales llamadas cargas sinápticas muestran gran estabilidad en almacenar los conocimientos (López & Caicedo, 2009). Las redes neuronales recurrentes logran el comportamiento dinámico realimentando sus entradas con muestras anteriores de las salidas, el comportamiento dinámico de las redes recurrentes hace que sean una poderosa herramienta para simular e identificar sistemas dinámicos no lineales o complejos (Díaz, 2016; González et al., 2017).

Con estos antecedentes y con el objetivo de establecer una herramienta operacional para la gestión ambiental, se aplicarán estas teorías para analizar la variación y predecir las concentraciones de PM2.5 en el Centro Histórico de la ciudad de Quito.

## Metodología

### *Teoría del caos*

El desarrollo de la teoría del caos se dio hacia finales del siglo XIX, cuando el físico–matemático Henri Poincaré introdujo el concepto de no linealidad, que contradecía el ideal clásico de los sistemas lineales, en los que causa y efecto se identifican plenamente y la resultante de un fenómeno es igual a la suma de sus partes. La no linealidad implica divergencia entre resultado y origen, y la simple adición de las partes de un hecho no corresponde al producto final, por lo que los algoritmos lineales poco pueden hacer para explicar su dinámica. El problema no era menor, si se considera que un 90% de los procesos de la naturaleza son de carácter no lineal (Motte & Campbell, 2013).

Para la aplicación de la teoría del caos se sigue un procedimiento que va desde la reconstrucción de las series de datos al espacio de las fases, la determinación del grado de casticidad que tiene el sistema y llegando a la predicción de dichos sistemas, para lo cual se usa el software Tisian (Hegger et al., 1999).

### *Tiempo de retardo (delay time) y dimensión de encaje (embedding dimension)*

El tiempo de retardo, es el paso más importante en el análisis del espacio de fases, es la técnica de reconstrucción del sistema, formando vectores y ejecutando una proyección en el espacio de fases, que es un espacio físico que permite detectar variables dinámicas, en este método se busca la relación biunívoca de los datos en el espacio de las fases (Ivancevic & Ivancevic, 2007).

La técnica de reconstrucción de espacio de fase más importante es el método de retardo. Los vectores en un nuevo espacio, el espacio de empotramiento, se forman a partir de los valores retardados de las mediciones escalares expresado en la ec.1:

$$S_n = (S_{n-(m-1)\tau}, S_{n-(m-2)\tau}, \dots, S_n) \quad (1)$$

El número m de elementos se denomina dimensión de encaje, el tiempo  $\tau$  se denomina generalmente retardo. Los teoremas incorporados por Takens y Sauer (Abderrahim et al., 2014), afirman que, si la serie  $s_n$  consiste en mediciones escalares del estado de un sistema dinámico, entonces bajo ciertas suposiciones, la incorporación de tiempo de retardo proporciona una imagen uno a uno del conjunto original, siempre que m sea lo suficientemente grande.

Si  $N$  mediciones disponibles, el número de vectores de encaje se estima que será  $N-(m-1) \tau$ . Esto debe tenerse en cuenta para la correcta normalización de las cantidades promediadas. Existe literatura sobre la opción "óptima" de los parámetros de encaje  $m$  (dimensión de encaje) y  $\tau$  (tiempo de retardo).

Fraser y Swinney el método de información mutua para calcular el retraso o llamado también tiempo de retardo con resultados razonables, tiene en cuenta las correlaciones no lineales, que se puede calcular con la Ec. 2 (Ivancevic & Ivancevic, 2007):

$$S = - \sum_{ij} P_{ij}(\tau) \ln \left( \frac{P_{ij}}{p_i p_j} \right) \quad (2)$$

Dónde para una partición en los números reales  $p_i$  es la probabilidad de encontrar un valor de serie temporal en el intervalo  $i$ -ésimo, y  $P_{ij}$  es la probabilidad conjunta de que una observación caiga en el intervalo  $i$ -ésimo y la posterior observación un tiempo  $\tau$ , caiga en el  $j$ -ésimo intervalo. En teoría, esta expresión no tiene dependencia sistemática del tamaño de los elementos de la partición y se puede calcular fácilmente. Existen buenos argumentos de que, si la información mutua exhibe un mínimo marcado en un cierto valor, entonces este es un buen candidato para un retraso razonable (Valentin, 2009).

La idea del algoritmo del falso vecino es que para cada punto  $\mathbf{s}_i$  en la serie de tiempo busque su vecino más cercano  $\mathbf{s}_j$  en un espacio  $m$ -dimensional. Por lo tanto, calcula la distancia  $\|\mathbf{s}_i - \mathbf{s}_j\|$ , que reincide en ambos puntos. Como se expresa en la ec. 3.

$$r_i = \frac{|s_{i+1} - s_{j+1}|}{\|\mathbf{s}_i - \mathbf{s}_j\|} \quad (3)$$

Al exceder  $s_j$  un umbral determinado  $s_j$ , este punto está marcado como un falso vecino. El criterio de que la dimensión de encaje es lo suficientemente alta, es la fracción de puntos para la cual  $s_i > s_j$  es cero, o al menos suficientemente pequeño. La introducción del concepto de falso vecino más cercano permite especificar una mínima separación temporal de los vecinos válidos y así determinar las dimensiones apropiadas para las matrices creadas, buscando vecinos a un punto fijo y que estén contenidos dentro de una bola de radio centrada en ese punto, de esta forma se determina la distancia interactiva y se calcula. Si la relación de la distancia de la interacción y del vecino más cercano excede a un valor determinado, el punto se marca como un vecino equivocado (Valentin, 2009).

### *Reducción del ruido*

El estudio de series de datos puede llevar a sacar conclusiones erróneas ya que los datos que se toman como magnitud de la variable a analizar suelen estar influidas por varios factores que introducen distorsiones (ruido) y por lo tanto no permiten observar con claridad los patrones dinámicos que se quiere analizar. Por lo tanto, es necesaria la

transformación de los datos de manera de poder separar el componente ruidoso y extraer señales que informen acerca del comportamiento evolutivo de la serie (Fernández, 2014).

Dentro del análisis de series temporales, algunos métodos para la reducción del ruido usan el hecho que los puntos, es decir los datos, están relacionados dinámicamente y que debido a hechos determinísticos los puntos en una misma región del espacio tienen un comportamiento similar. El algoritmo no lineal más simple sobre reducción de ruido consiste en reemplazar la coordenada central de cada vector de encaje, por un promedio local de esta coordenada (Pino-Vallejo et al., 2018). Como se expresa en la Ec. (4):

$$S_{n-\frac{m}{2}} = \frac{1}{U_e^n} \sum_{S_n \in U_e^n} S'_{n-\frac{m}{2}} \quad (4)$$

Al escoger una  $m$  dimensión de encaje y un adecuado tiempo de retardo  $\tau$ , por cada vector de encaje  $S_n$  se forma una vecindad  $U_{ne}$ , se forma en el espacio de fase que contiene todos los puntos  $S_n$  tal que  $\|S_n - S'\| < \epsilon$ , donde  $\epsilon$  es el radio de la vecindad, que debe ser lo suficientemente grande para cubrir la extensión de ruido y  $S'$  el espacio de fase inicial (Pino-Vallejo et al., 2018). Por cada vector de encaje  $S_n = (S_{n-(m-1)}, \dots, S_n)$  con  $\lambda=1$ , la coordenada corregida es  $S_{n-(m/2)}$ , se calcula como el promedio sobre la vecindad  $U_{ne}$ .

#### *Exponentes de Lyapunov*

El cálculo del espectro completo de Lyapunov requiere considerablemente más esfuerzo que el exponente máximo (Bochi & Rams, 2016). El proceso consiste en una estimación de los Jacobianos locales, es decir, de la dinámica linealizada que gobierna el crecimiento de las perturbaciones infinitesimales. El caos surge del crecimiento exponencial de las perturbaciones infinitesimales es así como para garantizar el análisis de esta inestabilidad están los exponentes de Lyapunov, que cuantifican el momento que dos trayectorias se alejan, como se expresa en la Ec 5.

$$\lambda = \frac{1}{t} \ln \left( \frac{r(t)}{r} \right) \quad (5)$$

Donde:  $r(t)$  es el semi-eje mayor de una elipse,  $r$  es el radio pequeño y  $t$  es el tiempo grande, en general se establece  $\lambda$  conocido como coeficiente de Lyapunov. En un sistema caótico al menos un exponente de Lyapunov es positivo.

El cálculo del espectro Lyapunov requiere considerablemente más esfuerzo que sólo el exponente máximo. Un método muy usado es la estimación de los jacobinos locales, es decir, de la dinámica linealizada, que gobierna el crecimiento de perturbaciones infinitesimales. Se encuentra a partir de ajustes directos de modelos lineales locales del tipo  $\delta_{n+1} = a_n \delta_n + b_n$ , de modo que la primera fila del jacobino es el vector  $a_n$ , y  $(J)_{ij} =$

$\delta_{i-1,j}$  para  $i=2,\dots,m$  donde  $m$  es la dimensión de incrustación o encaje. El  $a_n$  se halla con el método de los mínimos cuadrados  $\sigma^2 = \sum_l (\delta_{l+1} - a_n \delta_l - b_n)^2$  donde  $\delta_l$  es el conjunto de vecinos de  $\delta_n$  (Pino-Vallejo et al., 2018).

La Entropía de Kolmogórov-Sinai, también llamada métrica ( $h$ ), mide la pérdida de información a lo largo de la evolución del sistema. Como se expresa en la Ec. 6, este tipo de entropía es determinada por la suma de exponentes positivos de Lyapunov. Tiene importante significado cuando se aplica a sistemas de los cuales no se dispone más que de series temporales de valores de cierta variable (Pino-Vallejo et al., 2018).

$$h = \sum_{\lambda_i > 0} \lambda_i \quad (6)$$

Este tipo de entropía se relaciona con la previsibilidad que expresa una escala de tiempo promedio hasta la cual las predicciones pueden ser posibles. Esta cantidad invariante intrínseca del sistema tiene una relevancia singular para la caracterización de la previsibilidad. Los valores más altos de  $h$  implican valores más bajos de la predictibilidad  $1/h$ , es decir, cuanto más caótico, menos predecible (Domański & Ławryńczuk, 2017).

También de estos coeficientes se puede determinar la dimensión fractal usando el método conocido de Kaplan – Yorke, 1979, Ec. 7 (Medvinsky et al., 2017; Stolz & Keller, 2017).

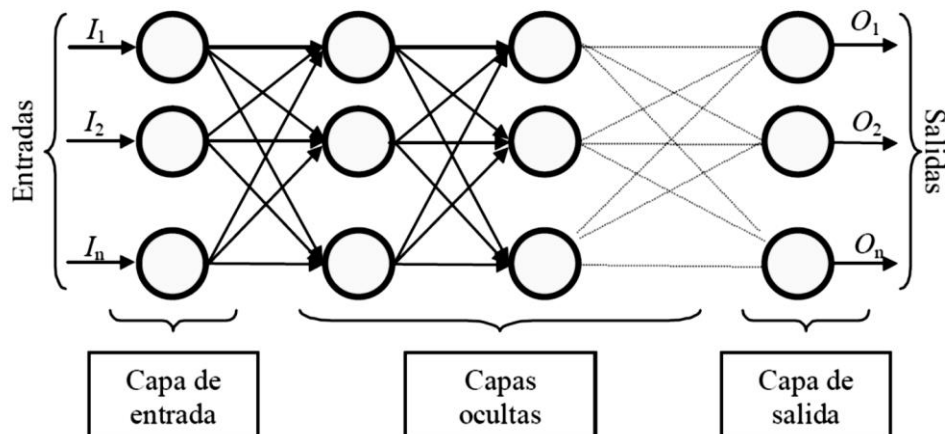
$$D_F = j + \frac{\sum_{i=1}^j \lambda_i}{|\lambda_{j+1}|}, \text{ con } \sum_{i=1}^j \lambda_i < 0 \quad (7)$$

### *Redes Neuronales*

Una red neuronal está construida por un conjunto de nodos interconectados que se denominan neuronas y se organizan en capas. Una red neuronal básica tiene una capa de entrada (input layer), una capa intermedia oculta (hidden layer) y una capa de salida (output layer). Los nodos de la capa de entrada corresponden con el número de variables que queremos introducir en la red (serían los equivalentes a las variables independientes en una regresión lineal). Por otro lado, el número de nodos de salida depende del número de variables a predecir. Por último, la capa intermedia se usa principalmente para aplicar transformaciones no lineales a las variables de entrada originales (Kocak et al., 2020).

**Figura 1**

*Ejemplo de una red neuronal totalmente conectada*



**Fuente:** Lewis (2016)

Las redes neuronales dado a su naturaleza y fundamentos poseen un gran número de rasgos similares a las neuronas del cerebro. Tienen la facultad de instruirse de la experiencia, como de sistematizar sucesos pasados y para poder llevarlos a nuevos sucesos (Cruz et al., 2007).

#### *Redes Neuronales Recurrentes*

Las redes neuronales recurrentes (Recurrent Neural Networks) tienen una gran habilidad para aprender problemas difíciles que tratan sobre datos con series temporales. Estas redes tienen estados ocultos distribuidos en el tiempo que permiten almacenar mucha información sobre el pasado de forma eficiente (Géron, 2017).

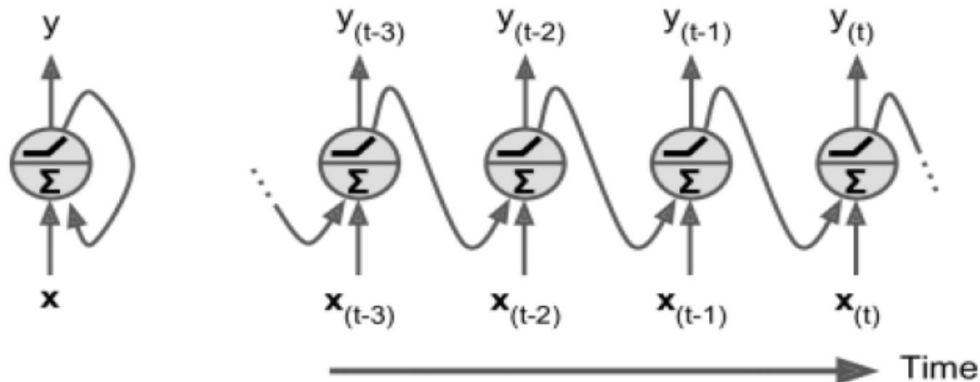
Las redes recurrentes son dinámicas, es decir también tienen conexiones hacia atrás. La salida depende de los actuales valores de las variables de entrada a la red, de los anteriores valores de la entrada, de los anteriores valores de la salida, etc.

Fijémonos en la más sencilla red neuronal recurrente posible, compuesta por sólo una neurona que recibe los inputs y que produce un output y que este lo envía a sí misma como vemos en la parte izquierda de la siguiente imagen:



**Figura 2**

*Neurona recurrente (izquierda), desarrollo en el tiempo (derecha)*



**Fuente:** Lewis (2016)

Cada neurona recurrente tiene dos conjuntos de pesos; uno para los inputs  $x(t)$  y otro para las salidas del paso temporal anterior  $y(t-1)$ . Llamemos a estos pesos  $w_x$  y  $w_y$ . La salida  $y(t)$  de una neurona recurrente se puede calcular de la siguiente manera, Ec. 9:

$$y(t) = \varphi(x_{(t)}^T \cdot w_x + y_{(t-1)}^T \cdot w_y + b) \quad (9)$$

donde  $b$  es el sesgo (bias) y  $\varphi$  es la función de activación. Podemos resumir la expresión anterior mediante el álgebra lineal de la siguiente manera, Ec. 10:

$$y(t) = \varphi(Xx_{(t)} \cdot W_x + Y_{(t-1)} \cdot W_y + b) = \varphi(|X_{(t)} Y_{(t-1)}|. W + b) \quad (10)$$

siendo  $W = [W_x \ W_y]^T$  y donde:

$Y(t)$  es una matriz  $m \times n$  neuronas que contiene los outputs del marco temporal  $t$  para cada batch (que es el conjunto de datos que entran a la neurona) en este caso  $m$  y  $n$  neuronas es el número de neuronas de la capa.

$X(t)$  es una matriz  $m \times n$  inputs que contiene las variables inputs de las  $m$  observaciones y  $n$  inputs es el número de variables explicativas de entrada.

$W_x$  es una matriz  $n$  inputs  $\times$   $n$  neuronas conteniendo los pesos de las conexiones para las entradas del paso temporal actual.

$W_y$  es una matriz  $n$  neuronas  $\times$   $n$  neuronas conteniendo los pesos de las conexiones para las salidas del paso temporal anterior.

Las matrices de pesos  $W_x$  y  $W_y$  se pueden juntar en la matriz  $W$  que será  $(n \text{ inputs} + n \text{ neuronas}) \times n \text{ neuronas}$

$b$  es un vector de tamaño neuronas que permite añadir un término constante (bias) al modelo.

Hay que darse cuenta de que  $Y(t)$  es una función de  $X(t)$  y de  $Y(t-1)$ , quien a su vez es función de  $X(t-1)$  y de  $Y(t-2)$ , quien es función de  $X(t-2)$  y así sucesivamente. Esto hace que  $Y(t)$  sea función de todas las variables de entrada desde el tiempo inicial  $t=0$  (esto es, de  $X(0); X(1); \dots; X(t)$ ). En  $t=0$  no hay outputs anteriores luego se asumen como ceros.

#### *Redes Neuronales Recurrentes de Elman y Jordan*

Las Redes Neuronales Recurrentes son adecuadas para modelar series temporales, se trabajó con redes Elman y Jordan, el análisis se realizó en el software libre R-Studio.

Una red neuronal de Elman es una red con tres capas (entrada, oculta y salida) a las que se le añade una capa con 'unidades de contexto' (o también llamadas unidades de retardo). Cada capa contiene una o más neuronas que propagan la información de una capa a la siguiente con la regla de aprendizaje que se haya fijado.

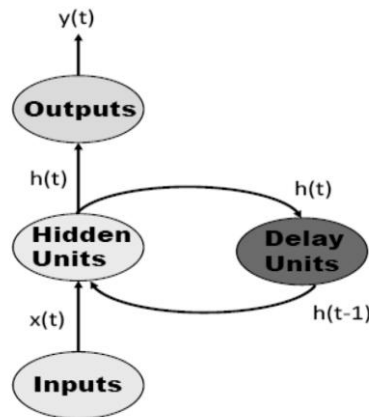
En las redes de Elman el número de neuronas de la capa de contexto tiene que ser igual al número de neuronas de la capa oculta. Además, todas las neuronas de la capa de contexto están conectadas con una neurona de la capa oculta.

El proceso de memoria se produce a través de las unidades de retardo que son alimentadas por las neuronas de la capa oculta. Los pesos de las conexiones entre la capa oculta y las unidades de retardo son fijos e iguales a 1. Gracias a esto en las unidades de retardo siempre mantenemos una copia de los valores de salida de las neuronas de la capa oculta del paso previo.

Podemos ver una representación de la red neuronal de Elman en la figura 3:

**Figura 3**

*Representaciones de la red neuronal de Elman*



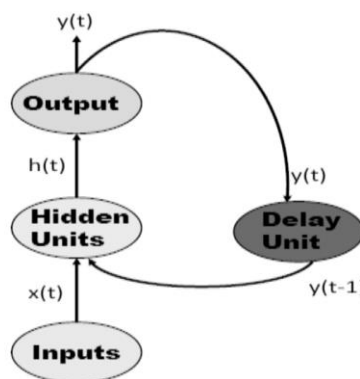
Fuente: Lewis (2016)

En la figura 3 se ve el esquema de la red de Elman, observando que la salida de la capa oculta se almacena en la capa de retardo, y que luego esta sirve de entrada en el siguiente paso para la capa oculta, donde las unidades de retardo  $u_i$  en las que se almacena el valor de la salida de la neurona  $y_i$  y sirven como entrada en el siguiente paso (Lewis, 2016).

La red neuronal de Jordan tiene una arquitectura similar a la de Elman, es decir, consta de la capa de entrada, una capa oculta, la capa de salida y la capa de contexto. La única diferencia es que ahora se guarda la información de la capa de salida en vez de la información de la capa oculta, por lo que la capa de contexto está conectada a la capa oculta y a la de salida como vemos en la figura 4. Por tanto, esta red recuerda la salida final del paso anterior en vez de la salida de la capa oculta. La estructura de la red neuronal de Jordan es la siguiente:

**Figura 4**

*Representaciones de la red neuronal de Elman*



Fuente: Lewis (2016)

Las ecuaciones matemáticas que definen ambos procesos son Ec. 11 y 12:

Red de Elman:

$$\begin{aligned}h_t &= \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \\y_t &= \sigma_h(W_y h_t + b_y)\end{aligned}\tag{11}$$

Red Jordan:

$$\begin{aligned}h_t &= \sigma_h(W_h x_t + U_h y_{t-1} + b_h) \\y_t &= \sigma_h(W_y h_t + b_y)\end{aligned}\tag{12}$$

Donde  $x_t$  es el vector con las variables de entrada,  $h_t$  es el vector de la capa oculta,  $y_t$  es el vector de salida,  $W = [W_x \ W_y]^T$  y  $U_h$  son las matrices de pesos y  $b = [b_h \ b_y]^T$  el bias.  $\sigma_h$  y  $\sigma_y$  son las funciones de activación de las respectivas capas.

Para la presente técnica no es obligatorio describir supuestos como la distribución de probabilidad, patrón de comportamiento de los datos de la serie para pronosticar de manera eficiente, además al ser utilizada para realizar predicciones presenta una característica significativa y es que los datos no serán desarrollados para probar supuestos como tendencia o estacionalidad en la serie previo a la realización de los pronósticos (Li et al., 2017).

Es necesario realizar un cambio de escala a los datos originales debido a que los valores de la ecuación logística se encuentran en el rango de (0,1), se aplica la siguiente formula:

$$y_b = \frac{x_i - x_{min}}{x_{max} - x_{min}}\tag{13}$$

Se ha establecido un ritmo de aprendizaje de 0.07 (Zhou et al., 2017), con un número máximo de 5000 iteraciones.

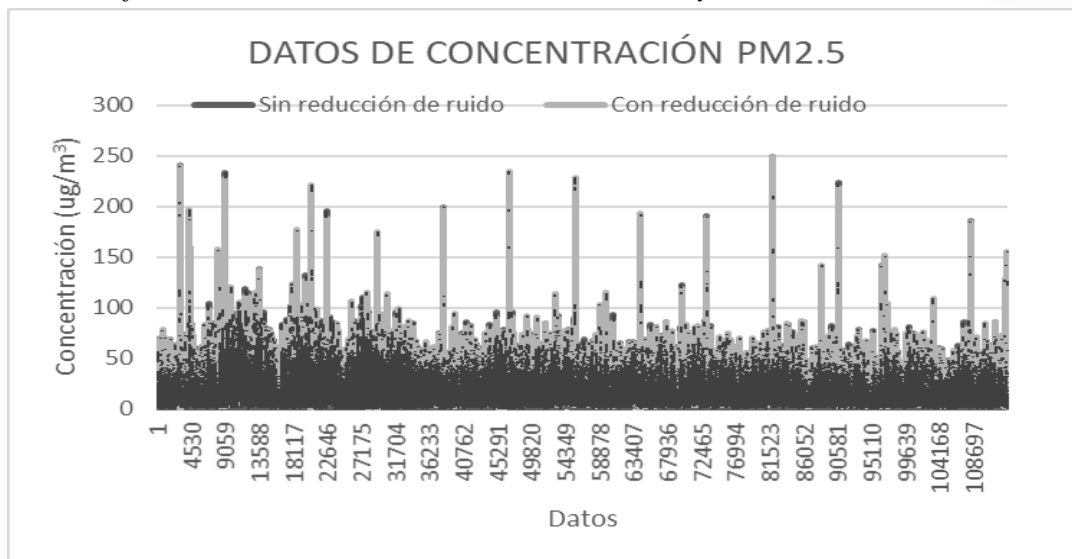
El uso de la Teoría del Caos ayuda a establecer el comportamiento dinámico del sistema y establecer su regularidad en el espacio de las fases, además se reduce el ruido en la serie datos, los cuales se normaliza y se aplica redes neuronales recurrentes.

### Resultados y discusión

Con los datos recolectados se procedió a desarrollar el análisis, en primera instancia una reducción de ruido en la serie de datos, con los que se trabajó para establecer la dinámica del sistema, procediéndose antes a normalizar los datos, y finalmente se predijo con redes neuronales recurrentes, para lo cual se ha usado el modelo TISIAN y el paquete R estudio.

**Figura 5**

*Gráfico de datos PM2.5 con reducción de ruido y sin reducción del ruido*

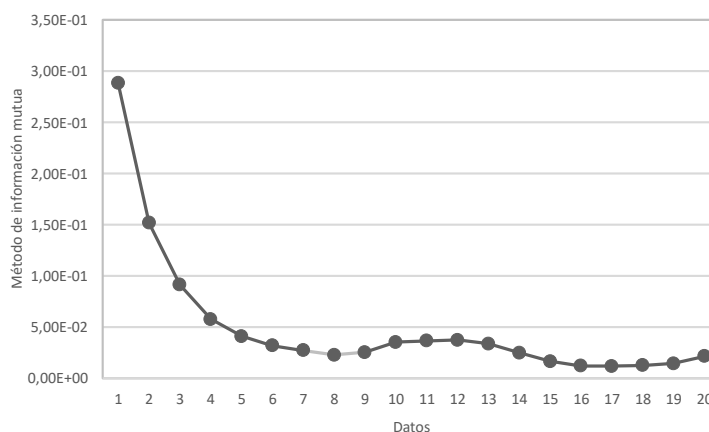


En la figura 5 se observa los datos con reducción del ruido y sin reducción de ruido, donde no se puede notar diferencias significativas entre las series, procedimiento necesario para evitar anomalías entre los datos y asegurar su correlación en el espacio de las fases.

En principio para realizar el análisis se normalizó los datos entre cero y uno, para mantener la variación de estos según los requerimientos de las redes neuronales recurrentes, verificándose que aporta al análisis usando la teoría del caos, ya que los rangos de variación no son muy grandes, además, por la cantidad de datos usados el sistema se muestra estable en los parámetros calculados.

**Figura 6**

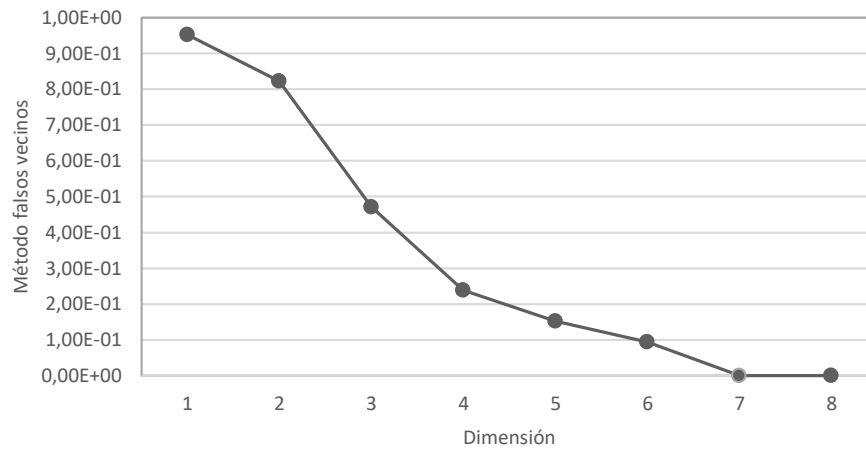
*Estimación del tiempo de retardo usando el método de información mutua*



En la figura 6 se presentan los cálculos aplicando el método de información mutua para hallar el tiempo de retardo, para lo cual se recomienda elegir el primer mínimo de la función.

**Figura 7**

*Estimación de la dimensión de encaje usando el método de falsos vecinos*



En la figura 7 se observa los resultados del cálculo de la dimensión de encaje usando el método de los falsos vecinos, el cual toma como valor más probable el primer cero que alcanza la función.

**Tabla 1**

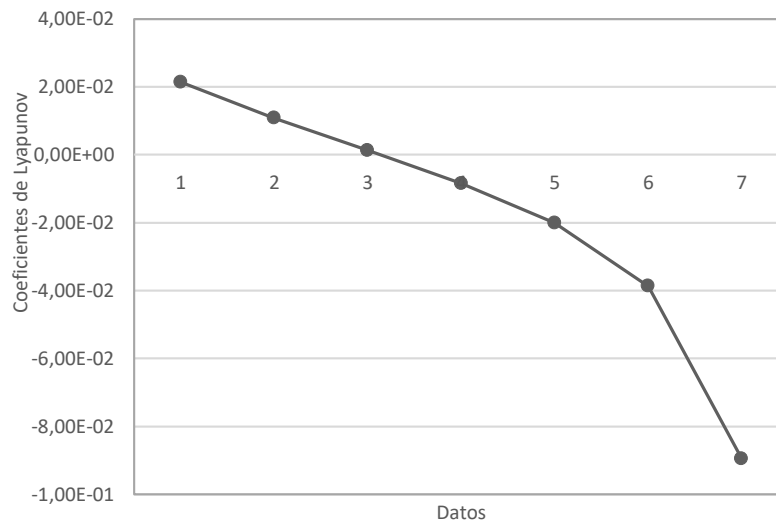
*Tiempo de retardo y dimensión de encaje con reducción del ruido*

Parámetros	Tiempo de Retardo	Dimensión de Encaje
PM 2.5	8	7

En la tabla 1 se observa el tiempo de retardo para la reconstrucción de los datos en el espacio de las fases y la dimensión de encaje que establece la dimensión mínima que se debe usar para analizar los datos observados.

**Figura 8**

*Gráfico de datos los coeficientes de Lyapunov calculados*



La figura 8 presenta los coeficientes de Lyapunov, indican la dinámica del sistema, cuando existe al menos un coeficiente positivo, se dice que el sistema es caótico, es decir son datos no lineales, en los cuales es conveniente aplicar métodos propuestos, Teoría del Caos y Redes Neuronales.

**Tabla 2**

*Dimensión fractal de Kaplan-Yorke y entropía de Kolmogoróv-Sinai*

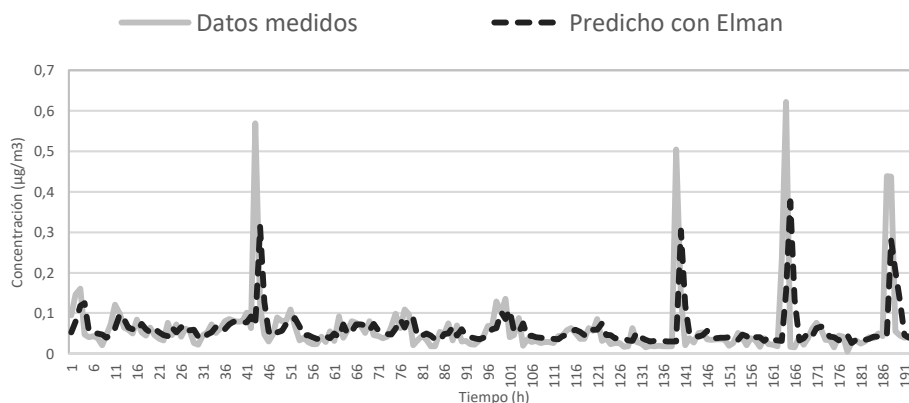
Parámetros	Dimensión Fractal	Entropía
PM2.5	5,12	0,0321

La tabla 2 muestra la dimensión de Kaplan-Yorke y la Entropía de Kolmogórov-Sinai, del sistema que identifica la dimensión fractal de este y el grado de desorden que los caracteriza.

Se determinó que la dimensión es fraccionaria, es decir es de tipo fractal, de 5,12 que ratifica la complejidad del sistema, de igual manera se aprecia la entropía del sistema que mide el grado de desorden del sistema y está asociada a su predictibilidad.

**Figura 9**

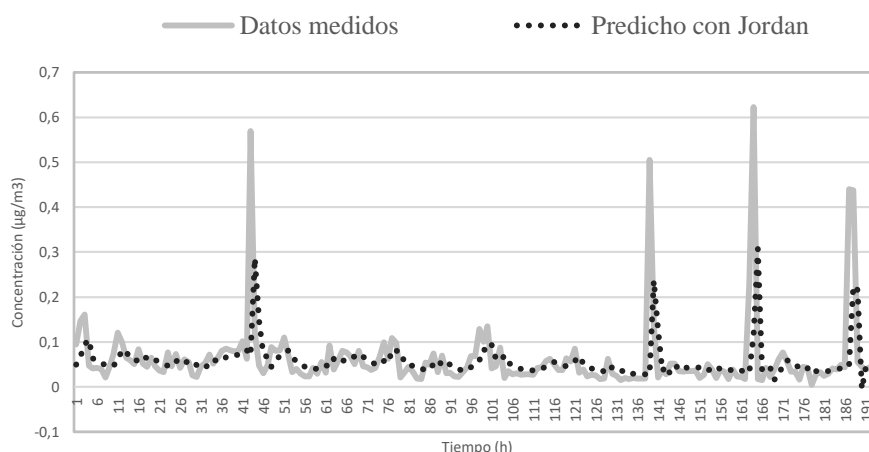
*Gráfico de datos PM2.5 medidos y predichos con redes neuronales recurrentes de Elman*



En la figura 9 se observa la comparación entre los datos reales y predichos usando redes neuronales recurrentes de Elman, tomando como base la dimensión de 7 determinada con la teoría del caos en la cual se describirían los datos con un error de 0,07 y 5000 repeticiones en el entrenamiento de las redes, se puede notar una tendencia muy buena entre los datos reales y datos predichos.

**Figura 10**

*Datos PM2.5 medidos y predichos con redes neuronales recurrentes de Jordan*



En la figura 10 se observa la comparación entre los datos reales y predichos usando redes neuronales recurrentes de Jordan, tomando como base la dimensión de 7 determinada con la teoría del caos en la cual se describen los datos con un error de 0,07 y 5000 repeticiones



en el entrenamiento de las redes, similar a lo aplicado en Elman, de igual forma se puede notar una tendencia muy buena entre los datos reales y datos predichos.

**Tabla 3**

*Análisis de varianza de los datos predichos*

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos				0,461	0,631	3,011
Dentro de los grupos						
Total	1,917	581				

También se aplica un análisis de varianza en la tabla 3, para determinar si existe diferencia entre las series de datos tanto predichas con redes neuronales de Elman, como de Jordan, obteniéndose que efectivamente no existe diferencia entre estas series.

**Tabla 4**

*Error cuadrático medio entre series de datos reales con predichos con la teoría del caos y Redes Neuronales Recurrentes*

Método	Error cuadrático medio	Error cuadrático medio porcentual respecto al rango de variación.
Método de Elman	0,08062	4,98
Método de Jordan	0,08003	4,94

Finalmente se determinó el error cuadrático medio entre los datos predichos y los datos medidos (tabla 4), y se calculó el error porcentual que se daría en el rango de variación de la serie medida, para tener una idea de la precisión del cálculo llegando en ambos casos alrededor de 4,9 %, demostrando que cualquiera de ellos podría ser aplicado en este sistema.

### Conclusiones

- Se aplica teoría del caos a las series de datos normalizados con reducción de ruido, obteniéndose resultados estables en los parámetros de análisis de la Teoría del Caos, con una dinámica caótica (2 coeficientes positivos de Lyapunov), que indica la irregularidad de los datos.
- Se establece la entropía de Kolmorov-Siani (0,0321) y la dimensión fractal de Kaplan-Yorke (5.2), resultados que confirman la complejidad del sistema y la dificultad en su predictibilidad.

- Al aplicar la dimensión de encaje como variable de entrenamiento de las redes se establece, que los dos métodos propuestos no presentar diferencias significativas al 0.05 de significancia, y su error porcentual medio cuadrático respecto al rango de variación de los datos de alrededor del 5 %.
- El conocimiento de la dinámica del sistema ayuda a preestablecer el comportamiento de los datos antes de predecir con redes neuronales recurrentes, dando información valiosa para saber las potenciales resultados.

### Agradecimiento

Al Proyecto de Monitoreo Hídrico en la Provincia de Chimborazo del Centro de Energías Alternativas y Ambiente (CEAA).

### Referencias Bibliográficas

- Abderrahim, N. W., Benmansour, F. Z., & Seddiki, O. (2014). A chaotic stream cipher based on symbolic dynamic description and synchronization. *Nonlinear Dynamics*, 78(1), 197–207. <https://doi.org/10.1007/s11071-014-1432-z>
- Bochi, J., & Rams, M. (2016). The entropy of Lyapunov-optimizing measures of some matrix cocycles. *Journal of Modern Dynamics*, 10, 255–286. <https://doi.org/10.3934/jmd.2016.10.255>
- Cruz, B., Martínez, S., Abed, R., Ábalo, G., Lorenzo, G., Matilde, M., & Lorenzo, M. G. (2007). Redes neuronales recurrentes para el análisis de secuencias Recurrente neurales / network for sequences analysis. *Rcci*, 1, 11.
- Díaz, M. (2016). Redes neuronales recurrentes y series temporales. <http://software-tecnico-libre.es/es/articulo-por-tema/todas-las-secciones/todos-los-temas/todos-los-articulos/redes-neuronales-recurrentes-y-series-de-tiempo>
- Domański, P. D., & Ławryńczuk, M. (2017). Assessment of predictive control performance using fractal measures. *Nonlinear Dynamics*, 89(2), 773–790. <https://doi.org/10.1007/s11071-017-3484-3>
- Fernández, D. G. (2014). Reducción del ruido y predicción de series temporales de alta frecuencia mediante sistemas dinámicos no lineales y técnicas neurales. *Banco Central Del Uruguay*, (1688), 15. <http://www.bcu.gub.uy/Estadisticas-e-Indicadores/Documentos de Trabajo/1.2014.pdf>
- Fowler, Andrew., & McGuinness, M. (2010). *Chaos: An Introduction for Applied Mathematicians*. (S. Nature, Ed.). New Zeland.

- [https://books.google.com.ec/books?hl=es&lr=&id=SdzODwAAQBAJ&oi=fnd&pg=PR8&dq=Chaos+An+Introduction+for+Applied+Mathematicians,+Switzerland,+Springer&ots=rqAReTU6eR&sig=TPdrhtsZG9OCH0j--37iAzdPIc#v=onepage&q=Chaos An Introduction for Applied Mathematician](https://books.google.com.ec/books?hl=es&lr=&id=SdzODwAAQBAJ&oi=fnd&pg=PR8&dq=Chaos+An+Introduction+for+Applied+Mathematicians,+Switzerland,+Springer&ots=rqAReTU6eR&sig=TPdrhtsZG9OCH0j--37iAzdPIc#v=onepage&q=Chaos+An+Introduction+for+Applied+Mathematician)
- Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn & TensorFlow*. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781491962282/>
- González, J. C., Tudurí, J. M., & Rul-lan, G. (2017). Análisis de Series Temporales Usando Redes Neuronales Recurrentes. <https://www.apsl.net/blog/2017/06/14/analisis-de-series-temporales-usando-redes-neuronales-recurrentes/>
- Hegger, R., Kantz, H., & Schreiber, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos*, 9(2), 413–435. <https://doi.org/10.1063/1.166424>
- Ivancevic, V. G., & Ivancevic, T. T. (2007). Introduction to Attractors and Chaos. In *High-Dimensional Chaotic and Attractor Systems* (pp. 1–151). Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-1-4020-5456-3\\_1](https://doi.org/10.1007/978-1-4020-5456-3_1)
- Kocak, C., Dalar, A. Z., Cagcag Yolcu, O., Bas, E., & Egrioglu, E. (2020). A new fuzzy time series method based on an ARMA-type recurrent Pi-Sigma artificial neural network. *Soft Computing*, 24(11), 8243–8252. <https://doi.org/10.1007/s00500-019-04506-1>
- Lewis, N. (2016). *Deep time series forecasting with python*. (C. I. P. Platform, Ed.). Retrieved from <https://www.goodreads.com/book/show/33843624-deep-time-series-forecasting-with-python>
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, 231, 997–1004. <https://doi.org/10.1016/j.envpol.2017.08.114>
- López, Jesús., & Caicedo, E. (2009). *Una aproximación práctica a las Redes Neuronales Artificiales*. (P. E. U. del Valle, Ed.) (Primera). Universidad del Valle. [https://www.researchgate.net/publication/303365431\\_Una\\_aproximacion\\_practica\\_a\\_las\\_Redес\\_Neuronales\\_Artificiales](https://www.researchgate.net/publication/303365431_Una_aproximacion_practica_a_las_Redес_Neuronales_Artificiales)
- Medvinsky, A. B., Nurieva, N. I., Rusakov, A. V., & Adamovich, B. V. (2017). Deterministic chaos and the problem of predictability in population dynamics. *Biophysics (Russian Federation)*, 62(1), 92–108. <https://doi.org/10.1134/S0006350917010122>

- Motter, A. E., & Campbell, D. K. (2013). Chaos at fifty. *Physics Today*, 66(5), 27–33. <https://doi.org/10.1063/PT.3.1977>
- Pino-Vallejo, M., Tierra, A., Haro, A., & Perugachi, N. (2018). Prediction of concentrations of PM2.5 in downtown Quito using the chaos theory. *AIP Conference Proceedings*, 2003, 1–8. <https://doi.org/10.1063/1.5050365>
- Rebala, G., Ravi, A., & Churiwala, S. (2019). *An Introduction to Machine Learning*. (Springer International Publishing, Ed.). 1. [https://books.google.com.ec/books/about/An\\_Introduction\\_to\\_Machine\\_Learning.html?id=gckwygEACAAJ&redir\\_esc=y](https://books.google.com.ec/books/about/An_Introduction_to_Machine_Learning.html?id=gckwygEACAAJ&redir_esc=y)
- Rodríguez, A. (2015). *Contaminación atmosférica y justicia ambiental en Quito*. FLACSO. <https://repositorio.flacsoandes.edu.ec/xmlui/handle/10469/8549>
- Stolz, I., & Keller, K. (2017). A general symbolic approach to Kolmogorov-Sinai entropy. *Entropy*, 19(12), 1–19. <https://doi.org/10.3390/e19120675>
- Valentin, A. (2009). *Y bolotin a tur, V. Yanovsky, chaos, concept, control and consnhuctive Use*. (Caron, Ed.) (Springer C). Springer. [https://www.researchgate.net/publication/281089929\\_YBolotinATurVYanovskyChaosConceptcontrol\\_and\\_Consnhuctive\\_UseSpringerISBN\\_978-3-642-00936-5](https://www.researchgate.net/publication/281089929_YBolotinATurVYanovskyChaosConceptcontrol_and_Consnhuctive_UseSpringerISBN_978-3-642-00936-5)
- Zhou, S., Li, W., & Qiao, J. (2017). Prediction of PM2.5 concentration based on recurrent fuzzy neural network. *Chinese Control Conference, CCC*, 3920–3924. <https://doi.org/10.23919/ChiCC.2017.8027970>

El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital**.



#### Indexaciones

