

Machine Learning y Caos

DANIEL MONTESINOS CAPACETE¹, DOLORES ESTEVE DÍAZ²

¹ Grado en Física, Universidad de Granada.

² Doble Grado en Física y Matemáticas, Universidad de Granada.

24 de junio de 2023

En este trabajo se estudia la relación entre Caos y Machine Learning mediante el uso de una red neuronal tipo *Echo State Network* (ESN) para predecir series caóticas clásicas. Mediante los atractores de Lorenz, Rössler y Mackey-Glass se estudia el concepto de caos y la definición e importancia de los exponentes de Lyapunov para luego exportar estos conocimientos al análisis de la dinámica interna de la ESN. Se demuestra cómo a ESN alcanza un pico óptimo de predicción en el umbral crítico de su dinámica de activaciones. Además, se analiza el efecto de implementar reglas de plasticidad sináptica en la dinámica del reservorio. Estos estudios se acompañan adicionalmente con un análisis de recurrencia que permite visualizar gráficamente la dinámica de la red.

ÍNDICE

1. Introducción	1
2. Fundamentos de caos	2
2.1. Exponentes de Lyapunov locales y globales	2
2.2. Sistemas caóticos clásicos	2
2.3. Estudio del MLE para el sistema Mackey-Glass	2
3. Echo State Networks	3
3.1. Modelo básico	4
3.2. Echo State Property (ESP)	4
3.3. Cálculo del MLE de las activaciones	4
3.4. Reglas de plasticidad	5
3.5. Representational drift	5
4. Análisis de Recurrencia	5
5. Resultados	6
5.1. Simulación	6
5.2. Predicción de la ESN: Punto Crítico y rendimiento óptimo	7
5.3. Plasticidad anti-Hebbiana	9
6. Conclusiones	12
Referencias	13

1. INTRODUCCIÓN

El estudio del caos dentro de la rama de los Sistemas Complejos es fundamental para entender diversos fenómenos que aparecen en la naturaleza.

Una dinámica es caótica cuando pequeñas variaciones en las condiciones iniciales pueden llevar a resultados drásticamente distintos. Esta sensibilidad afecta directamente a la capacidad de predicción en muchos de estos sistemas. Esta impredecibilidad se produce a pesar de ser sistemas completamente deterministas con ecuaciones conocidas.

Sin embargo, aunque pueda parecer aleatorio, estos sistemas exhiben patrones y estructuras universales y por tanto, susceptibles de estudio. La comprensión de estas características comunes nos puede ayudar a revelar la dinámica subyacente de un sistema complejo y proporcionar información valiosa sobre su comportamiento.

En este trabajo se hace uso de redes neuronales artificiales de tipo *Echo State Network* (ESN) para intentar predecir series temporales de atractores caóticos clásicos como el de Lorenz, Rössler o Mackey-Glass y estudiaremos el caos desde una perspectiva de Machine Learning.

Las ESN derivan de las *Recurrent Neural Networks* (RNN) y se incluyen en el campo de Reservoir Computing. Las RNN son el modelo más cercano a la dinámica del cerebro y por tanto, de la inteligencia [10]. Además, más allá de predecir series temporales, el objetivo será estudiar como cambia la dinámica de la ESN y buscar los hiperparámetros óptimos de la misma. Mediante el cálculo del exponente de Lyapunov global de las activaciones neuronales, se pretende encontrar dichos hiperparámetros óptimos analizando la zona cerca del punto crítico en la que dicho exponente de Lyapunov pasa a ser positivo.

Por último, se analiza cómo cambia la dinámica inter-

na de la red al introducir reglas de plasticidad sináptica que actualizan las conexiones entre neuronas. Estas reglas de plasticidad están inspiradas en observaciones biológicas como se detalla más adelante.

2. FUNDAMENTOS DE CAOS

Según Strogatz [17], se define el **caos** como comportamiento *aperiódico* en un sistema *determinista* muy sensible a las *condiciones iniciales*.

En detalle:

1. *Comportamiento aperiódico* significa que hay trayectorias que no convergen a un punto fijo ni a una órbita periódica ni casi periódica cuando $t \rightarrow \infty$.
2. *Determinista* significa que no hay parámetros o inputs aleatorios o ruidosos. La irregularidad del comportamiento viene dada por la no linealidad del sistema.
3. *Sensible a las condiciones iniciales* significa que trayectorias inicialmente cercanas se separan exponencialmente rápido (exponente de Lyapunov positivo).

2.1. Exponentes de Lyapunov locales y globales

La dinámica dentro de un atractor es muy sensible a las condiciones iniciales y en consecuencia, dos trayectorias inicialmente cercanas divergen rápidamente de forma exponencial. Esto significa que la predicción de un sistema de este estilo es muy complicada debido a que las incertidumbres se amplifican a gran velocidad [17].

El espectro de Lyapunov es el conjunto de coeficientes que determina el comportamiento real del sistema. Se define a partir de considerar la evolución de una n -esfera infinitesimal [18]. Esta esfera se convierte en un elipsoide y los coeficientes se calculan en términos de la longitud de los ejes principales ($p_i(t)$) en relación a las condiciones iniciales ($p_i(0)$).

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t} \log_2 \frac{p_i(t)}{p_0(t)} \quad \text{con } \lambda_i > \lambda_j \forall i < j \quad (2.1)$$

Además, cualquier sistema que contenga al menos un exponente de Lyapunov positivo se define como caótico con la magnitud de este exponente reflejando la escala temporal para la cual la dinámica del sistema se convierte en impredecible. Según [17], el tiempo de predicción se puede expresar como

$$t_{\text{horizon}} \sim O\left(\frac{1}{\lambda} \ln \frac{a}{\|\delta_0\|}\right) \quad (2.2)$$

con a un factor numérico.

El diámetro del elipsoide está controlado por el máximo exponente de Lyapunov (MLE). Si definimos $\delta(t)$ como la diferencia entre dos trayectorias, se observa que

$$\|\delta(t)\| \sim \|\delta_0\| e^{\lambda t}.$$

Por ejemplo, para el atractor de Lorenz se obtiene lo siguiente:

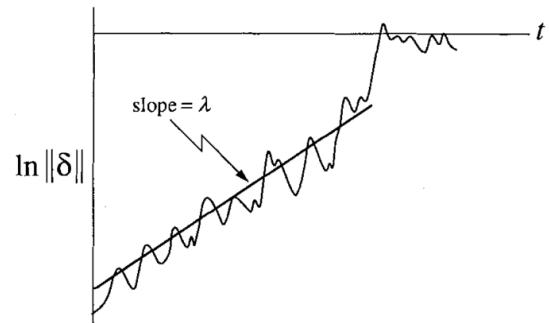


Fig. 2.1. Cálculo del MLE para el atractor de Lorenz [17].

Hay que tener en cuenta que la curva no es idealmente recta y las fluctuaciones se deben a que la fuerza de la divergencia exponencial varía a lo largo del atractor. Además, la divergencia se detiene cuando la separación es comparable al diámetro del atractor. Las trayectorias no pueden alejarse entre sí más que eso y en consecuencia la curva se satura.

Sistema	Lorenz	Rössler	Mackey-Glass
MLE	1.5	0.09	0.0074

Tabla 2.1. Tabla resumen del MLE teórico de los sistemas clásicos [13].

2.2. Sistemas caóticos clásicos

Se presenta en primer lugar una tabla resumen Tab. 2.2 donde se pueden consultar las ecuaciones correspondientes a cada modelo así como los parámetros estándar que se suelen manejar y que se utilizarán a lo largo del trabajo.

Sistema	Ecuaciones	Parámetros	$\Delta t(s)$	$\vec{r}(0)$
Lorenz	$\dot{x} = \sigma(y - x)$	$\sigma = 10$		$x(0) = 1$
	$\dot{y} = x(\rho - z) - y$	$\rho = 28$	0.01	$y(0) = 0.5$
	$\dot{z} = xy - \beta z$	$\beta = 2.67$		$z(0) = 0.9$
Rössler	$\dot{x} = -y - z$	$a = 0.2$		$x(0) = -1$
	$\dot{y} = x + ay$	$b = 0.2$	0.1	$y(0) = 1$
	$\dot{z} = b + z(x - c)$	$c = 0.7$		$z(0) = 1$
M-G	$\dot{x} = \frac{ax_\tau}{1-x_\tau^c} - bx$	$a = 0.1 \quad b = 0.2$		
		$c = 10 \quad \tau = 17$	1	$x(0) = 0.3$

Tabla 2.2. Ecuaciones y parámetros estándar de los modelos estudiados [17].

Modelo de Lorenz

Introducido por Edward Lorenz, en 1963, es un sistema dinámico determinista tridimensional no lineal deri-

vado de las ecuaciones simplificadas de rollos de convección que se producen en las ecuaciones dinámicas de la atmósfera terrestre. Las corrientes de convección son las corrientes de aire que se forman debido a la tendencia ascendente del aire caliente y descendente del aire frío. Un rollo de convección es la forma que toma una nube debido a estas corrientes.

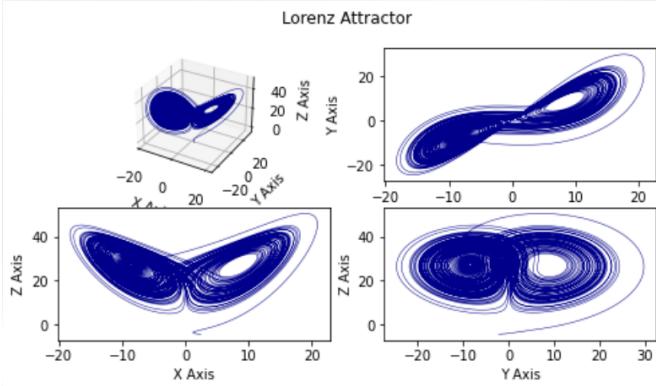


Fig. 2.2. Simulación del atractor de Lorenz en las condiciones estándar Tab. 2.2.

Este es probablemente el atractor más famoso ya que el concepto de *efecto mariposa* surgió a raíz de su forma y caló en el imaginario colectivo (Fig. 2.2).

Rössler

Otto Rössler diseñó el atractor Rössler Fig. 2.3 en 1976 con el propósito de encontrar un sistema que se comportara como el de Lorenz, pero que fuera más fácil de estudiar. Posteriormente se utilizó para modelar el equilibrio de reacciones químicas como la famosa reacción de Belousov-Zhabotinsky que se estudia también en la asignatura de *Biofísica*.

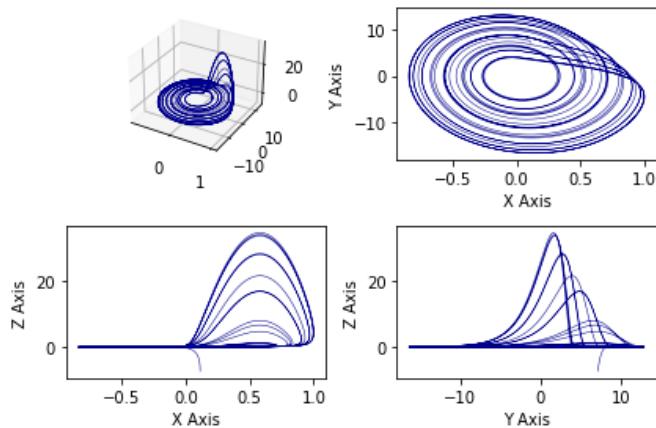


Fig. 2.3. Simulación del atractor de Rössler en las condiciones estándar (2.2).

Mackey-Glass

Este modelo fue introducido por Michael Mackey y Leon Glass en 1977 para describir fenómenos sanos y

patológicos en determinados sistemas biológicos Fig. 2.4. Hay numerosas enfermedades serias cuyo principal síntoma es la alteración de la periodicidad o *delay* de un observable, $x_\tau = x(t - \tau)$. Por ejemplo, en estudios de control de la respiración o de la hematopoyesis se han asociado procesos patológicos con inestabilidades oscilatorias a modelos complejos [11].

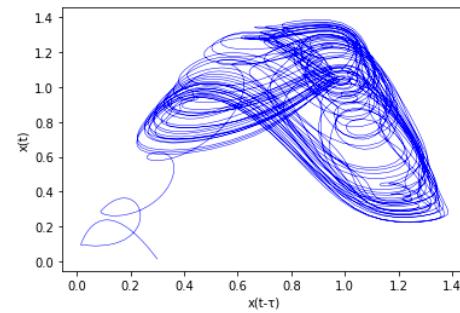


Fig. 2.4. Simulación del atractor de Mackey-Glass en las condiciones estándar Tab. 2.2 para $\tau = 30$.

2.3. Estudio del MLE para el sistema Mackey-Glass

Se fundamenta el cálculo del MLE de una serie temporal correspondiente al atractor de Mackey-Glass en el trabajo de E. Mendes [13]. Para este cometido se sugiere generar dos series interválicas a partir de la misma condición inicial cambiando de alguna forma el orden de las operaciones matemáticas en la ecuación diferencial Tab. 2.2. Además, con el objetivo de comparar resultados con [13], $\tau = 30$. En este caso, al no disponer de la suficiente precisión computacional para realizar ese método y se ha optado por usar la misma ecuación diferencial pero variando levemente la condición inicial.

Sea $x_0 = 0.3$ y tres condiciones iniciales ligeramente distintas tal que $x_1 = x_0 + 10^{-13}$, $x_2 = x_0 + 1.5 \cdot 10^{-13}$ y $x_3 = x_0 + 2 \cdot 10^{-13}$.

Se ha obtenido un valor promedio de

$$\lambda = (6.018 \pm 0.011) \cdot 10^{-3}.$$

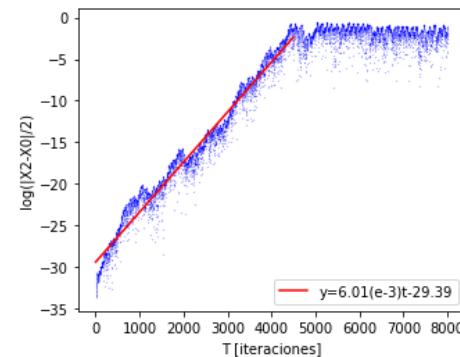


Fig. 2.5. Ajuste para x_2 a modo ilustrativo.

Lo más interesante de este resultados es que, no sólo se puede calcular el MLE, sino que además se verifica el fenómeno de saturación de la curva de ajuste. Se ve como en Fig. 2.5 sigue una curva consistente con lo que predecía Strogatz en Fig. 2.1 para el atractor de Lorenz. De hecho, en este caso, se reproducen bastante bien los resultados de [13].

Se ha preferido añadir este pequeño cálculo dentro del fundamento teórico ya que el grueso del proyecto es el estudio de la relación del caos con la dinámica interna de la ESN y su capacidad de predicción. De hecho, posteriormente se comentarán las correcciones en el cálculo del MLE dentro de la ESN.

3. ECHO STATE NETWORKS

3.1. Modelo básico

En este trabajo se utiliza un modelo básico de Echo State Network (ESN) al que posteriormente se añaden reglas de plasticidad. La ESN es un tipo de red neuronal recurrente constituida por tres capas principales: la capa de entrada, la capa interna y la capa de salida [10].

Dado un conjunto de puntos $\{u(t_i)\} \in \mathbb{R}$ correspondientes a una serie temporal caótica, estos se introducen como input en la red a través de la matriz $W^{in} \in \mathbb{R}^{n_x \times 2}$ que caracteriza la capa de entrada e incluye el sesgo (*bias*). La capa interna se define a partir de la matriz de conexiones neuronales $W^{res} \in \mathbb{R}^{n_x \times n_x}$ siendo n_x el número total de neuronas en la red. Las activaciones neuronales a cada tiempo t_i están definidas por un vector $\mathbf{x}(t_i) \in \mathbb{R}^{n_x \times 1}$ y se actualizan como

$$\mathbf{x}(t_i) = \tanh\left(W^{res}\mathbf{x}(t_{i-1}) + \varepsilon W^{in}[1; u(t_i)]\right), \quad (3.1)$$

donde ε es un factor de escala para el *input* de la red. La función de activación neuronal es la tangente hiperbólica y se encarga de introducir la no linealidad en el reservorio, lo que es fundamental para aumentar la capacidad de aprendizaje de la red. Finalmente, el output $y(t_i)$ se obtiene multiplicando el vector $\mathbf{x}(t_i)$ por la matriz $W^{out} \in \mathbb{R}^{1 \times n_x}$. Esto se escribe como

$$y(t_i) = W^{out} \mathbf{x}(t_i), \quad (3.2)$$

de forma que la matriz W^{out} constituye la capa de salida de la ESN. Dado un punto $u(t_i)$ de la serie temporal, el objetivo es que la red sea capaz de predecir el siguiente punto $u(t_{i+1})$ (*target*). Para lograr esto, primero se entrena la red de manera supervisada usando N puntos de la serie temporal. La ESN debe reproducir los datos de entrenamiento y también nuevas series temporales. Para ello, se emplea el método de regresión Ridge que ayuda a disminuir los problemas de sobreajuste introduciendo un parámetro de regularización β a la regresión lineal [10]. Por tanto, la matriz de salida se obtiene como

$$W^{out} = \mathbf{y}^{target} M(MM^T + \beta I)^{-1}, \quad (3.3)$$

donde $\mathbf{y}^{target} \in \mathbb{R}^{1 \times N}$, $M \in \mathbb{R}^{(2+n_x) \times N}$ e I es la matriz identidad. M es la matriz que surge de concatenar $[1; u(t_i), \mathbf{x}(t_i)]$ para cada uno de los N pasos temporales del entrenamiento. Las matrices W^{in} y W^{res} , se inicializan con valores aleatorios y se mantienen fijas en el modelo básico.

Los hiperparámetros más importantes que se estudian son la esparsidad o *sparsity*, el radio espectral de la matriz de conexiones del reservorio y el *input scaling* [10]. La *sparsity* define el porcentaje de ceros que tiene una matriz de conexiones interneuronales. Normalmente, se fijan a valores altos para obtener mejores resultados. Se generan con la misma *sparsity* tanto W^{in} como W^{res} y se fija este parámetro.

El radio espectral es el mayor autovalor de una matriz. En este caso, $\rho(W^{res})$ lo que hace es escalar la amplitud de la distribución de los elementos no nulos. Como se ve en la próxima sección, está directamente relacionado con la *Echo State Property*.

El *input scaling* se refiere a la escala de la matriz W^{in} , y se modifica cambiando el valor del factor de escala ε . Junto con el radio espectral, son los parámetros que van a determinar de forma drástica el comportamiento de la ESN y permitirán su optimización.

3.2. Echo State Property (ESP)

En teoría, para que una ESN pretenda funcionar, el reservorio debe satisfacer la condición de ESP. Esta propiedad se refiere a que el estado del reservorio debe de estar definido únicamente por algunos pasos previos del *input*. Es decir, para un *input* lo suficientemente largo, el estado del reservorio no debería depender de las condiciones iniciales. Altos valores del radio espectral $\rho(W^{res})$ pueden llevar a situaciones con múltiples puntos fijos, periódicas o incluso caóticas, violando esta propiedad. Por lo tanto, un radio espectral tal que

$$\rho(W^{res}) < 1, \quad (3.4)$$

asegura la ESP en una gran mayoría de situaciones.

La ESP sirve para simplificar el proceso de *training* en comparación con otro tipo de redes, ya que se fijan las sinapsis de forma aleatoria en un inicio y no se vuelven a modificar (sin plasticidad). Además, se supone que esta memoria que la red desarrolla le ayuda a enfrentarse al ruido y generalizar mejor, proporcionando así un rendimiento superior en la predicción.

Sin embargo, como la no linealidad de las funciones de tipo sigmoide utilizadas en la ESN fuerzan a la norma del vector de estados del reservorio a estar acotada, hay que revisar la condición (3.4). Se encuentra que la ESP puede cumplirse para radios considerablemente grandes [4, 10]. En conclusión, se pueden determinar valores óptimos de radio espectral que sean mucho más grandes que la unidad.

3.3. Cálculo del MLE de las activaciones

Para posteriormente estudiar la criticalidad del sistema, lo que se va a hacer es calcular el MLE de las activaciones neuronales. Sin embargo, es muy recomendable alterar escrupulosamente el método de cálculo del MLE para evitar *overflows*, es decir, números que exceden el rango permitido y no se pueden representar con precisión. El procedimiento que se sigue dentro del código es el siguiente [5]:

1. Se perturban n neuronas de una red y la otra no se altera. Entonces, el estado de la red perturbada \mathbf{x}^2 se separa de la inalterada \mathbf{x}^1 en una cantidad γ_0 .
2. Se va avanzando la simulación y se recogen las diferencias k -ésimas como $\|\mathbf{x}^1(k) - \mathbf{x}^2(k)\|$. En este caso se usa la norma euclídea pero se pueden utilizar otras.
3. Se resetea el estado de la red perturbada como $\mathbf{x}^1(k) + \frac{\gamma_0}{\gamma_k}(\mathbf{x}^2 - \mathbf{x}^1)$ como se ve en la figura (3.1). Esta renormalización es precisamente el detalle que evita el *overflow* mencionado.

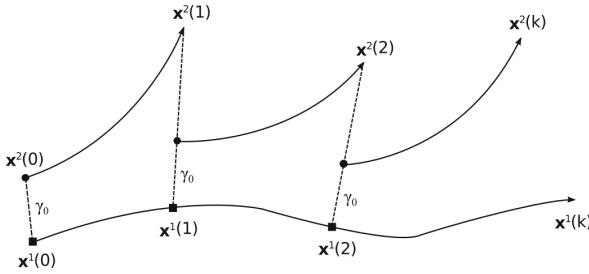


Fig. 3.1. Estimación numérica del MLE con las correcciones de [5].

3.4. Reglas de plasticidad

El modelo de ESN discutido en la Sec. 3.1 mantiene fija la matriz de conexiones neuronales que se genera aleatoriamente. Sin embargo, es natural preguntarse si existe alguna forma de elegir o cambiar los pesos de dicha matriz de manera que se mejore la predicción de la red. En este punto es donde entran en juego las reglas de plasticidad.

La plasticidad es un concepto que se utiliza en neurociencia para hacer referencia a la capacidad que tiene el cerebro de cambiar su estructura neuronal [3]. A grandes rasgos, hay dos tipos de reglas de plasticidad bien diferenciadas: plasticidad sináptica y plasticidad intrínseca. La plasticidad sináptica consiste en la capacidad de adaptación de las conexiones neuronales y nace de la idea de que "Neuronas que disparan juntas, se conectan juntas"(Donald Hebb, 1949 [8]). Por otro lado, la plasticidad sináptica consiste en la regulación de la

intensidad con la que disparan las neuronas.

Desde el punto de vista de la ESN, las reglas de plasticidad surgen como herramienta de aprendizaje no supervisado cuyo objetivo es entender y aprovechar los mecanismos neurológicos observados en el cerebro. La plasticidad sináptica consiste en ir actualizando los pesos de la matriz W^{res} mientras que la plasticidad intrínseca se centraría en modificar la Ec. (3.6) para cambiar como se activan las neuronas. En este trabajo, se implementa reglas de plasticidad sináptica exclusivamente. Concretamente, se utilizará la regla anti-Hebbiana normalizada que actualiza la matriz de pesos como [14]

$$W_{kj}^{res}(t_{i+1}) = \frac{W_{kj}^{res}(t_i) - \eta x_k(t_{i+1})x_j(t_i)}{\sqrt{\sum_l (W_{kl}^{res}(t_i) - \eta x_k(t_{i+1})x_l(t_i))^2}}, \quad (3.5)$$

donde se define η como el parámetro de aprendizaje y $x_j(t_i)$ corresponde a la activación de la neurona j del vector $\mathbf{x}(t_i)$. La normalización asegura que $\sum_j [W_{kj}^{res}(t_i)]^2 = 1$ lo que evita que las conexiones cambien de manera descontrolada. La Ec. (3.5) actualiza los pesos sinápticos de forma que neuronas que se activan juntas, disminuyen su conexión sináptica. Esto es contrario a la idea inicial de Hebb, sin embargo, existe evidencia experimental de plasticidad anti-Hebbiana [1, 9] y también hay estudios que sugieren que esta regla mejora la dinámica de las neuronas corticales [2].

3.5. Representational drift

El cerebro renueva continuamente sus componentes moleculares y celulares, y los sustratos neuronales de sus experiencias y memorias están sujetos a un cambio continuo [16]. Esta aparente inestabilidad desafía la idea de que la conectividad sináptica y las respuestas neuronales individuales se correlacionen directamente con la memoria. En [15] se sugiere añadir ruido al sistema (*representational drift*) y estudiar cómo varía la capacidad de predicción. Este fenómeno ocurre cuando los patrones de activación de las neuronas del reservorio o las conexiones sinápticas experimentan cambios significativos. Puede ser positivo en cuanto a que la red se adapte a cambios en el *input*, esto puede ser útil en tareas donde la entrada pueda tener cierta estocasticidad. Sin embargo, también puede ser perjudicial al perder la red la capacidad de generalización y conducir a un rendimiento deficiente en tareas de predicción.

En esta situación, se implementa tal y como se muestra en la Ec. 3.6:

$$\mathbf{x}(t_i) = \tanh \left(W^{res} \mathbf{x}(t_{i-1}) + \varepsilon W^{in}[1; u(t_i)] \right) + \xi \quad (3.6)$$

El término ξ es el término de ruido. Se pueden estudiar ruidos de distintos tipos, uniformes, gaussianos, correlados, etc, y se puede estudiar si cada uno de ellos afecta de manera distinta al comportamiento de la ESN.

4. ANÁLISIS DE RECURRENCIA

Una técnica utilizada en el análisis la dinámica de sistemas no lineales consiste en estudiar las recurrencias que aparecen en la correspondiente serie temporal. Una recurrencia hace referencia a la repetición de un estado o patrón en un sistema a lo largo del tiempo, es decir, un sistema muestra recurrencia cuando vuelve a un estado previamente visitado en su evolución dinámica. Esto se ve exemplificada en la Fig. 4.1. El estudio de las recurrencias permite ver cuando el sistema vuelve a pasar por el mismo punto del espacio de fases, lo que permite encontrar patrones ocultos en la dinámica del mismo. El análisis

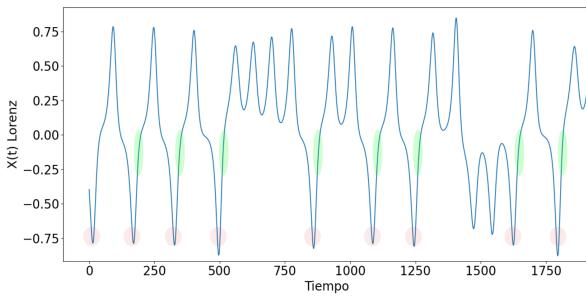


Fig. 4.1. Coordenada $X(t)$ del sistema de Lorenz generado con los parámetros de la tabla 2.2. Las partes sombreadas indican algunos ejemplos de recurrencia en la serie.

de recurrencia para el estudio de una serie temporal $\{u(t_i)\}_{i=1}^T$ se implementa mediante el cálculo de la matriz de recurrencias $R \in \mathbb{R}^{T \times T}$. Esta matriz se define como [4]

$$R_{ij} = \Theta(\tau_{RP} - d[u(t_i), u(t_j)]) \text{ con } i, j \in [1, T] \quad (4.1)$$

donde $d[\cdot, \cdot]$ es una medida de disimilitud sobre el espacio de fases, $\Theta(\cdot)$ es la función de Heaviside de forma que

$$\tau_{RP} > d[u(t_i), u(t_i)] \rightarrow R_{ii} = 1, \quad (4.2)$$

$$\tau_{RP} < d[u(t_i), u(t_i)] \rightarrow R_{ii} = 0, \quad (4.3)$$

y $\tau_{RP} > 0$ es un umbral para identificar las recurrencias. La elección de la medida de disimilitud $d[\cdot, \cdot]$ y del valor de τ_{RP} es libre y según el problema a estudiar. En este caso, la idea es obtener la matriz R para la serie temporal de los sistemas caóticos clásicos y de la actividad en la red neuronal. Siguiendo el trabajo realizado en [4], en este caso se utiliza la distancia de Manhattan como medida disimilitud. Para dos puntos de la serie temporal se calcula como

$$d[u(t_i), u(t_j)] = |u(t_i) - u(t_j)|, \quad (4.4)$$

y para el vector de activaciones neuronales $x(t_i)$ se tiene

$$d[x(t_i), x(t_j)] = \sum_{k=1}^{n_x} |x_k(t_i) - x_k(t_j)|. \quad (4.5)$$

A partir de la Ec. (4.1) es posible calcular R que es una matriz de ceros y unos indicando que puntos son recurrentes o no. Los elementos de matriz igual a uno serán aquellos cuya distancia de disimilitud es menor que el valor umbral τ_{RP} y son considerados recurrentes. En este trabajo se ha utilizado como valor umbral el 20 % de la distancia media entre todos los pasos temporales para cada caso. Es posible analizar la dinámica de la serie temporal y de la ESN a partir de sus respectivas matrices R de forma cuantitativa (Análisis de Cuantificación de Recurrencia) [4]. Sin embargo, no ha sido posible implementar estas técnicas ya que tienen un elevado coste computacional. Aún así, existe la posibilidad de realizar el plot de recurrencia (RP) que consiste en representar la matriz R en un mapa de color. El RP es una herramienta gráfica muy útil para analizar de forma cualitativa la dinámica de la serie temporal (o actividades neuronales) a partir de la visualización de los puntos recurrentes de la misma. El RP va a depender del tipo de la distancia de disimilitud y de la elección de τ_{RP} pero trabajando de forma consistente (usando siempre las mismas condiciones) será posible ver como cambia la dinámica en el reservorio en función de los parámetros de la ESN.

Como se discute en [12], una dinámica caótica se puede ver reflejada en el RP como la aparición de líneas diagonales paralelas a la diagonal principal (que siempre es uno por definición). Las líneas diagonales muestran los puntos separados temporalmente que pasan por la misma zona del espacio de fases lo que se puede relacionar con posibles atractores de la dinámica. De hecho, en [6] se discute que existe una relación (no inmediata) entre la longitud de las diagonales y el MLE.

5. RESULTADOS

5.1. Simulación

En este apartado se establecen algunos parámetros y procedimientos generales que se realizan en todas las simulaciones para la ESN. En primer lugar se genera una serie temporal $\{u(t_i)\}_{i=1}^{N+N'} \in \mathbb{R}$ de los sistemas clásicos vistos en la Sec. 2.2. Para todas las simulaciones, $N = 4000$ es el número de puntos de la serie que se usan para entrenamiento y los siguientes $N' = 1000$ se usan para el test de predicción. El número total de neuronas es $n_x = 300$, la sparsity se fija al 90 % y para la regresión lineal se usa $\beta = 10^{-7}$ en todos los casos. El scaling se mantiene a $\varepsilon = 1$ excepto cuando se indique lo contrario. El radio espectral $\rho_{sp} \equiv \rho(W^{res})$ no se fija a ningún valor concreto en general y mayoritariamente se estudia la dinámica de la red en función de este parámetro. Cuando se utilicen las reglas de plasticidad, la fase de entrenamiento consistirá en una o más epochs o épocas. Cada época corresponde a entrenar la red con los mismos $N = 4000$ puntos de entrenamiento. Todos los RC se obtienen a partir de las series temporales y actividades neuronales durante el periodo de entrenamiento. Tanto para Lorenz como para Rössler que son

sistemas tridimensionales, se toman series temporales correspondientes solo a la coordenada X.

En cada una de las simulaciones se realiza la fase de entrenamiento y posteriormente se realiza el test de N' puntos. Como la serie que corresponde al test es la continuación de la serie de entrenamiento, el test comienza con el estado final de activaciones neuronales de entrenamiento. Cuando esto ocurre, se obtiene un input de la red a partir de W^{out} en cada paso. Este *input*, se vuelve a introducir en la red y así se va generando la predicción.

Con el objetivo de cuantificar como de buena es la predicción se define el *Farthest Predicted Point* (FPP). Es el punto más lejano de la predicción que se considera correctamente predicho en base a una cota. En todas las simulaciones se toma como criterio que la diferencia entre la predicción y la serie real no sea mayor que el 3 % de la diferencia espacial entre los puntos más distanciados de la serie real. También se puede analizar la variación del Error Cuadrático Medio Normalizado (NRMSE). Sin embargo, una vez avanzado el proyecto, se pudo comprobar que, en comparación con el resto de parámetros analizados o incluso solamente con el FFP, no estaba proporcionando información valiosa. Muchas veces, que el NRMSE fuera bajo no implicaba ningún tipo de mejora en la predicción, por lo tanto, se ha optado por obviarlo en el comentario de los resultados. No quiere esto decir que este no sea un parámetro digno de estudio, sino que en este caso en concreto, no se han conseguido extraer conclusiones relacionadas.

5.2. Predicción de la ESN: Punto Crítico y rendimiento óptimo

El objetivo es estudiar como cambia la predicción de la ESN en función de los hiperparámetros ρ_{sp} y ε con la idea de encontrar un punto crítico en el que se mejoren los resultados de la misma. En primer lugar, para demostrar la existencia de esta criticalidad, se ha variado solamente el parámetro ρ_{sp} y se ha calculado tanto el MLE como el FPP en los tres sistemas descritos en la Sec. 2.2. Los datos obtenidos para estas simulaciones se pueden ver en la Fig. 5.1. En este caso, se ha repetido cada experimento 20 veces para obtener medidas más precisas y poder establecer barras de error en los mismos.

Se puede ver como la región en la que el MLE está transicionando de negativo a positivo se obtiene un pico en el valor del FPP y por lo tanto se verifica el efecto de esta criticalidad. Ocurre lo mismo para los tres sistemas dinámicos que se han introducido como input en la ESN. Por tanto, se verifica que la ESN mejora su predicción cerca del punto crítico. Estos resultados son parecidos a los que se pueden ver en [15].

Además de lo que se observa en la Fig. 5.1, el análisis de recurrencia apoya los resultados anteriores en referen-

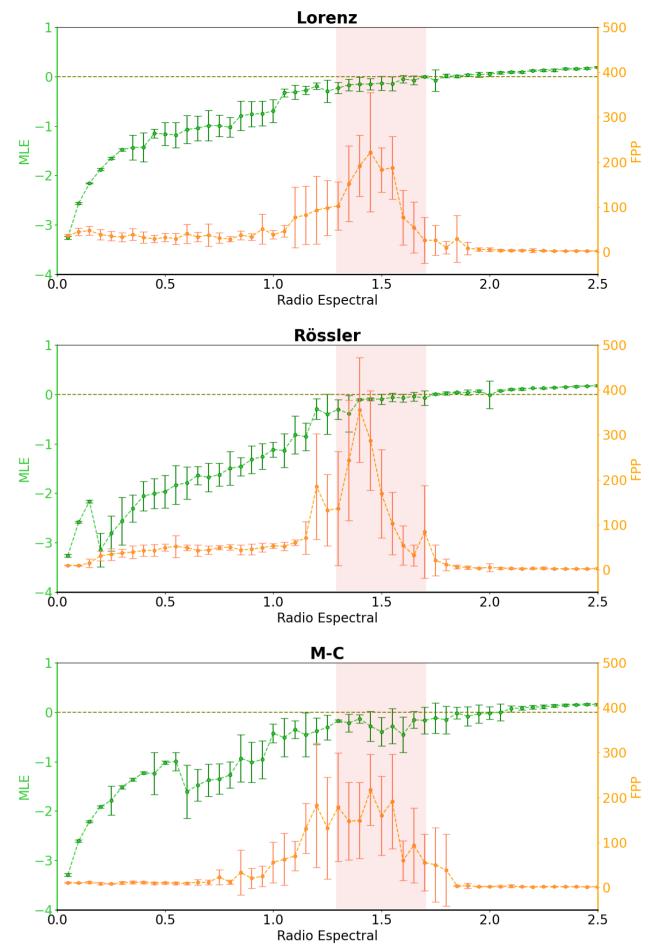


Fig. 5.1. MLE y FPP en función del radio espectral de la ESN. Se observa como en la región sombreada en rojo todas las predicciones mejoran y coinciden con el punto en el que la ESN se acerca al punto crítico.

cia a la posición del punto crítico respecto al valor del radio espectral. En la Fig. 5.2, se muestran varios RC correspondientes a las series temporales originales introducidas como *input* y a las activaciones neuronales durante el entrenamiento ($N = 4000$) según el radio espectral. Recordar que un punto blanco en el diagrama corresponde a un punto recurrente mientras que uno negro significa que no hay recurrencia. El RC se trata de hacer un plot de la matriz R que es una matriz de ceros y unos pero debido a la resolución de imagen no aparecen todos los píxeles y por eso hay tonos rojizos. En esta figura se han introducido los RC que se han considerado más representativos pero se ha realizado este tipo de plot para más valores del radio espectral. También es importante tener en cuenta que los RC mostrados en la Fig. 5.1 son una parte de plot total que sería de una matriz 4000×4000 . Esto se hace así ya que mostrando el RC completo no se aprecian algunos detalles.

El comportamiento que se observa en la Fig. 5.2 se describe a continuación. Cuando el radio espectral de la ESN

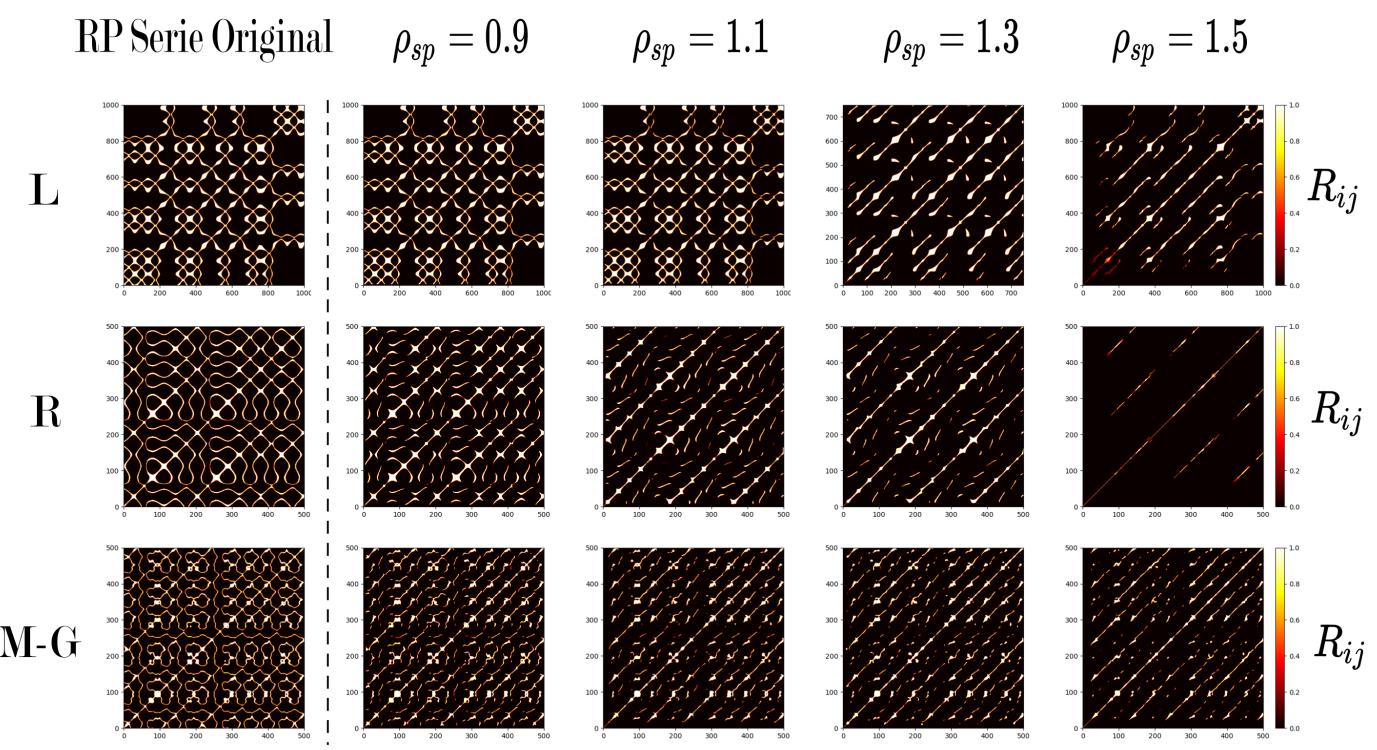


Fig. 5.2. RP para los tres sistemas clásicos en función del ρ_{sp} . Se observa como los patrones de la ESN se parecen a los de la serie original para valores del radio espectral bajos y cambian dando lugar a líneas diagonales.

es suficientemente bajo ($\rho_{sp} \approx 0.3 - 0.9$) se obtiene que el RC de las activaciones neuronales coincide exactamente con el RC que corresponde al input. Esto se puede ver para el caso del Lorenz con $\rho_{sp} = 0.9$ que corresponde a la primera fila de dicha figura. Para el Rössler y el M-G los patrones para este mismo radio son similares a la serie original pero con algunas deformaciones. Se observa que, al ir aumentando de forma progresiva el ρ_{sp} de la ESN, los patrones se van deformando y dan lugar a líneas diagonales paralelas. Si se aumenta el radio más allá de lo que se ve en la Fig. 5.2, todos los RC de la ESN terminan con un patrón parecido al que se ve para el caso del Rössler y $\rho_{sp} = 1.5$. De esta forma, al aumentar ρ_{sp} el movimiento de las activaciones neuronales de la ESN da lugar a un RC que correspondería a una dinámica caótica según lo discutido para series temporales en la Sec. 4.

Por tanto, los RC apoyan los resultados de la Fig. 5.1 en el sentido de que muestran que la transición en el punto crítico de la ESN ocurre para los mismos valores del radio espectral. Notar que el patrón que se obtiene en la Fig. 5.2 para el M-G coincide con el patrón obtenido en [4]. Para Rössler y Lorenz hay que tener en cuenta que el RC se calcula utilizando el input de la ESN que corresponde a la coordenada X de dichos sistemas.

Por otro lado, como ya se ha comentado, el otro hiperparámetro que va a modificar significativamente el comportamiento de la ESN es el ε (*scaling*). De esta forma, se ha calculado el FFP y el MLE como antes pero

barriendo valores de ρ_{sp} y ε con el objetivo de encontrar unos hiperparámetros que mejoren la predicción. El barrido se ha realizado en los intervalos $\rho_{sp} \in [0.95, 2.5]$ $\varepsilon \in [0.95, 2.5]$ y se ha repetido cada experimento 5 veces. El resultado de esto se puede ver en la Fig. 5.3 en el caso del atractor de Rössler. Observando dicha gráfica, se observa que existen unos parámetros para los que el FPP es máximo y está en torno a 600. Se consideran óptimos estos parámetros que maximizan el FPP dentro del rango de valores que se ha estudiado. En el gráfico se han omitido las barras de error para mayor claridad del mismo. Este análisis también se ha repetido para los otros dos sistemas clásicos pero no se muestra el gráfico para evitar ser redundantes. Los parámetros óptimos obtenidos para todos los sistemas junto con el mayor FFP calculado se recogen en la Tab. 5.1. En esta tabla se puede observar la magnitud de los errores obtenidos. Esto se debe a que por la propia variabilidad de la ESN, puede generar predicciones muy buenas o deficientes con los mismos parámetros. Esto es así ya que la matriz de entrada y la de conexiones internas se generan aleatoriamente. Es conveniente hacer cuantas más repeticiones posibles mejor para obtener más estadística y que los resultados sean más representativos. En este estudio se realizaron 5 repeticiones para cada caso con el objetivo de evitar tiempos de simulación excesivamente largos.

Es interesante destacar que los parámetros óptimos para cada sistema son diferentes mostrando que es necesaria

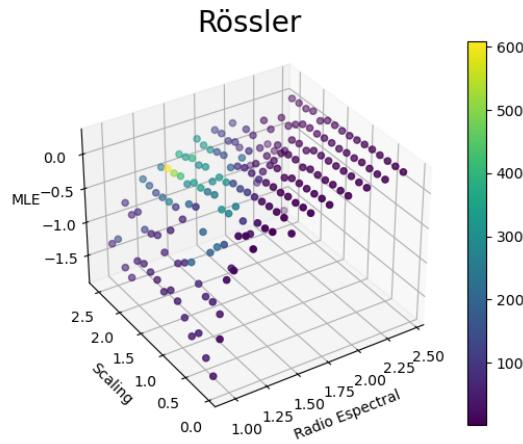


Fig. 5.3. Cálculo 3D del MLE en el eje Z y FPP en código de colores frente a ρ_{sp} y ϵ para el atractor de Rössler.

rio modificar la ESN dependiendo del problema que se pretenda estudiar con ella (Tab. 5.1). Por otro lado, es notable la diferencia del FPP óptimo en los diferentes casos siendo Rössler el sistema que mejor se ha podido predecir. Hay que tener en cuenta que cuanto mayor sea el MLE de un sistema caótico, más difícil será obtener una buena predicción del mismo. En ese sentido, es lógico que el peor resultado óptimo se haya encontrado para el sistema de Lorenz pero sería esperable que la predicción del M-G fuese mejor que la del Rössler (Tab. 2.1). Esto de puede deber a diferentes motivos. Es posible que en el rango de valores estudiados para los hiperparámetros no se encuentren los verdaderos hiperparámetros óptimos para los otros sistemas. Además, es posible que en las 5 repeticiones realizadas para cada experimento, la ESN no haya tenido "buena suerte" dando lugar a un FPP máximo más bajo del que realmente podría tener.

Sistema	ρ_{sp}	ϵ	MLE	FPP
Lorenz	1.25	0.5	-0.09 ± 0.03	170 ± 90
Rössler	1.4	2.4	-0.35 ± 0.02	600 ± 300
M-G	1.4	0.9	-0.34 ± 0.34	210 ± 50

Tabla 5.1. Parámetros óptimos calculados para la predicción de cada sistema caótico.

De forma ilustrativa, en la Fig. 5.4 se muestra la predicción asociada a los parámetros óptimos del Rössler (Tab. 5.1). En este caso, se ha obtenido un valor de $FPP = 648$. Como se ha comentado, no siempre que se realice la predicción usando los mismos parámetros se obtiene el mismo resultado. Para obtener este valor del FPP fue necesario hacer varias repeticiones en las que se obtenían valores en el rango de 100 – 500 hasta que saliera la predicción óptima.

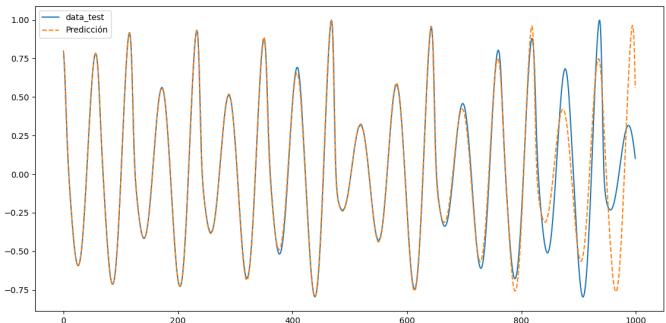


Fig. 5.4. Predicción para Rössler con los parámetros óptimos en Tab. 5.1 ($FPP = 648$).

5.3. Plasticidad anti-Hebbiana

Una vez estudiada la ESN básica, ahora se analiza la diferencia de comportamiento al añadir la regla de plasticidad anti-Hebbiana discutida en la Sec. 3.1. Para ello, se estudia la variación del FPP, el ρ_{sp} y el MLE en función de los *epochs* con los que se entrena la ESN usando dicha regla de plasticidad. La idea es entender como esta regla de plasticidad sináptica cambia la dinámica del reservorio. La regla de plasticidad cambia el ρ_{sp} ya que a cada paso del entrenamiento se actualizan los pesos de la matriz de conexiones sinápticas. Los hiperparámetros iniciales de la ESN en todos los casos son $\rho_{sp} = 0.95$ y $\epsilon = 1$ (este hiperparámetro no se ve afectado por la regla de plasticidad).

Se analiza la variación de los parámetros de interés para cada uno de los sistemas clásicos obteniendo el conjunto de gráficas que se ve en la Fig. 5.5. En esta figura se puede observar la variación del MLE y el radio espectral en función de las *epochs* así como el cálculo del FPP para cada caso. Para favorecer la estadística, cada *epoch* se ha repetido 5 veces. Hubiera sido deseable aumentar el número de repeticiones pero introducir la regla de plasticidad aumenta notablemente el tiempo de ejecución por lo que se ha limitado el número de repeticiones. Para estos resultados se ha usado $\eta = 10^{-6}$. Se observa que en todos los sistemas se tiene una tendencia inicial a aumentar la capacidad de predicción proporcionalmente al número de *epochs*, tal y como se veía en [14]. En el caso de M-G, se alcanza un pico de predicción en la *epoch* 8 y posteriormente se pierde considerablemente capacidad predictiva pero, debido al bajo número *epochs* que ha sido posible computar (10 *epochs*), no se puede asegurar que este número sea óptimo. Haciendo simulaciones sin calcular estadística (cada *epoch* solo se ejecuta una vez), fue posible observar que al aumentar el número de *epochs* el radio espectral seguía aumentando (recordar los parámetros óptimos anteriores Tab. 5.1). Sin embargo, no se observaban tendencias claras ya que al no hacer repeticiones sobre cada *epoch* no es posible cuantificar las desviaciones ni hacer medias. De todos modos, se observa como la regla de plasticidad cambia la matriz de conexiones interna de forma que el radio espectral aumenta y el FPP mejora con-

secuentemente para las primeras *epochs*. Además de esto, el MLE de la ESN también va variando y en general se mantiene cerca del punto crítico tomando valores dispersos entre 0 y -1 aproximadamente. Para el caso de M-G en la Fig. 5.5, se ve que al llegar al *epoch* 7 el MLE se sitúa ligeramente debajo del cero.

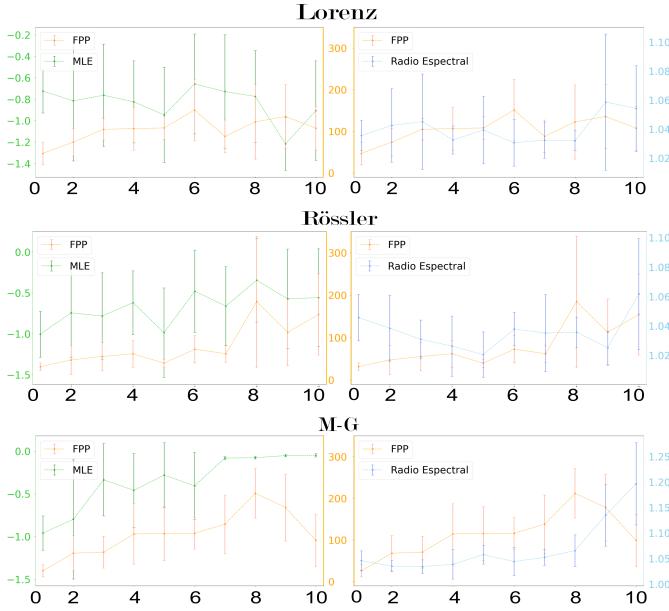


Fig. 5.5. Variación del ρ_{sp} , el MLE y el FPP frente al número de *epochs*.

Además de observar como varía ρ_{sp} , el MLE y el FPP, es interesante preguntarse que es lo que realmente ocurre internamente en el reservorio al añadir la regla de plasticidad. Para ello, se realiza un análisis adicional del *input* efectivo y la comparación dinámica entre con y sin plasticidad. A partir de la Ec. (3.6), se define el *input* efectivo de la neurona n como $\tilde{u}_n(t_i) = W_{0n}^{in} + W_{1n}^{in}u_n(t_i)$ [14]. La idea es analizar la dinámica de las neuronas en función de este *input* efectivo que se introduce de manera externa durante la fase de entrenamiento. Para ello, se realiza un *plot* de la activación de una neurona aleatoria en función de su *input* efectivo. Eso se realiza tanto para el caso de un entrenamiento con y sin plasticidad con el objetivo de ver la diferencia de la dinámica entre ambos casos. Las representaciones que se muestran a continuación se han obtenido usando el sistema del M-G como input para la ESN.

En la Fig. 5.6 se muestra un ejemplo de la dinámica de dos neuronas seleccionadas aleatoriamente en función de su *input* efectivo. Se observa como se enriquece la dinámica de las neuronas con plasticidad comparada con las neuronas del modelo básico. El área que ocupan las trayectorias de las neuronas con plasticidad es mucho mayor, lo que indica que estas neuronas recorren más el espacio de fases que tiene disponible y mejoran la dinámica del reservorio. Para este tipo de representación no se cal-

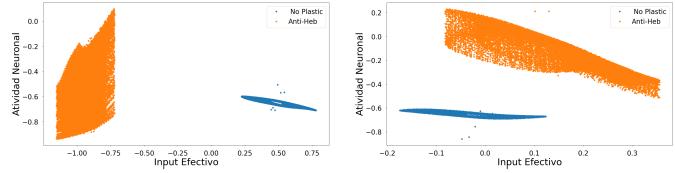


Fig. 5.6. Comparación de las trayectorias en el espacio de fases de dos neuronas distintas sin (azul) y con plasticidad (naranja) para 6 *epochs* y $\eta = 10^{-6}$.

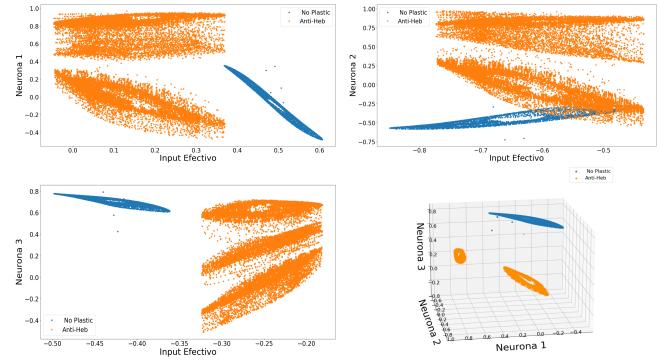


Fig. 5.7. Comparación de las trayectorias en el espacio de fases de tres neuronas distintas sin (azul) y con plasticidad (naranja) para 10 *epochs* y $\eta = 10^{-5}$.

culan ningún tipo de medias ya que se usan directamente las activaciones neuronales e *input* efectivo durante la fase de entrenamiento con los correspondientes *epochs*.

Es interesante estudiar lo que ocurre al aumentar mucho el número de *epochs* ya que según se sugiere en [15], un entrenamiento excesivo de la ESN usando la regla de plasticidad puede empeorar la predicción. Para estudiar esto, se vuelve a analizar la dinámica de las neuronas en función del *input* efectivo pero aumentando el parámetro de aprendizaje a $\eta = 10^{-5}$ con la idea de reducir el número de *epochs* necesarios. Se muestran los resultados en la Fig. 5.7 donde se ve que las neuronas con plasticidad cubren áreas aproximadamente disjuntas dentro del rango de posibles valores. Esto nos indica que dentro del reservorio aparece más de un atractor de la dinámica provocando que la actividad neuronal vaya alternando entre diferentes zonas del espacio de fases. Además, se ha representado en la parte inferior derecha de dicha figura la activación de las mismas tres neuronas en un espacio tridimensional. Se observa como la actividad de estas neuronas ocupa dos volúmenes separados del correspondiente espacio de fases.

Para lograr un entendimiento más profundo de lo que ocurre con la regla de plasticidad y corroborar los resultados obtenidos, se realiza un análisis de recurrencia sobre la actividad neuronal durante la última fase de entrenamiento para distintos *epochs*. Para estas simulaciones se utilizó $\eta = 10^{-6}$. Estos resultados se pueden observar en

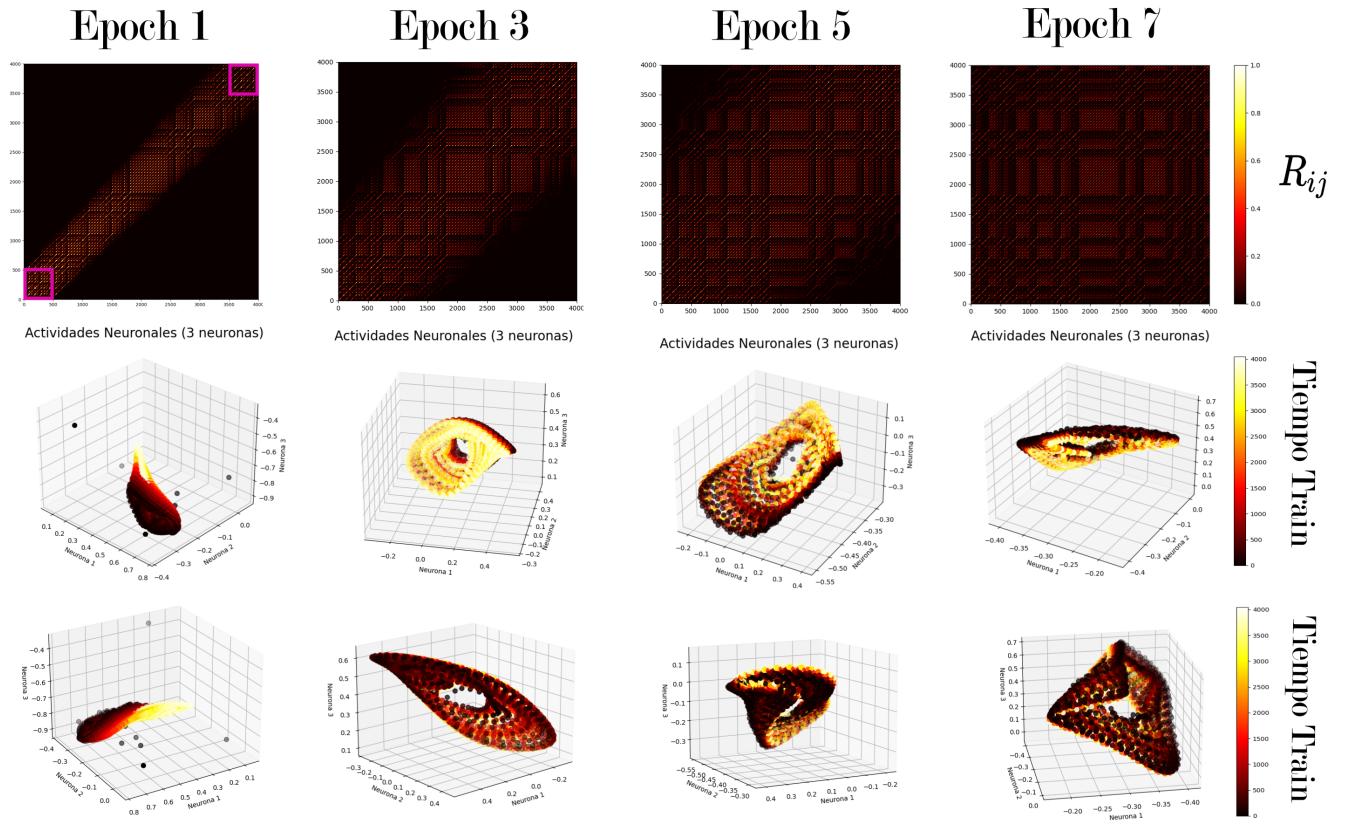


Fig. 5.8. RP para los tres sistemas clásicos con regla anti-hebbiana y radio espectral inicial $\rho_{sp} = 0.95$. Se muestra como cambia el RP y el movimiento de la actividad neuronal de tres neuronas en función de los epochs

la Fig. 5.8. En la parte superior de dicha figura se muestra el RP completo (4000×4000) para distintos epochs. Debajo de cada RP, se muestran dos planos diferentes de la misma representación tridimensional que corresponde a la actividad de tres neuronas aleatorias para cada epoch. El mapa de color de esta última indica el tiempo de la fase de entrenamiento para cada valor de las activaciones, es decir, los puntos oscuros son la actividad para el inicio de la fase de entrenamiento mientras que los más claros corresponden al final. Esto permite ver de forma gráfica como evoluciona conjuntamente la actividad de las neuronas en función de los pasos temporales de la fase de entrenamiento.

El RP que se obtiene para un número de epochs pequeño es bastante sorprendente. Se observa como se oscurece todo el RP excepto una diagonal bastante gruesa en la que se observan patrones habituales según lo visto en el apartado anterior (hay que hacer zoom en el RP para ello). Conforme aumenta el número de epochs, se va agrandando la diagonal central hasta que queda completamente cubierta. Para entender esto, hay que recordar que un punto oscuro indica que en el tiempo t_i y t_j la actividad neuronal del reservorio no está en el mismo punto del espacio de fases. Eso implica que, en los RP en los que se observa esta diagonal gruesa (epochs 1 y 3 de la

Fig. 5.8, principalmente), las actividades neuronales se mueven en el espacio de fases en una dirección concreta conforme avanza el entrenamiento. Esto último hace que no se repitan posiciones anteriores explicando las diagonales gruesas que aparecen en los RP. Este fenómeno se puede visualizar en las diferentes representaciones tridimensionales de la Fig. 5.8. Para una epoch, se observa como la activación neuronal evoluciona temporalmente de una zona oscura (inicio del entrenamiento) a una zona más clara (parte final del entrenamiento) sin repetir estados conforme avanza el entrenamiento. Al aumentar el número de epochs, esta evolución temporal no es tan clara pero se observa como no hay una dirección única de evolución y los diferentes puntos empiezan a mezclarse haciendo que el RP tenga un patrón completo. La regla de plasticidad durante el primer epoch conduce la actividad neuronal a diferentes regiones del espacio según la Ec. (3.5) y al aumentar los epochs se incrementa la extensión de la zona del espacio de fases en la que se mueven la actividad neuronal.

Además, en la Fig. 5.8 se han resaltado dos cuadrados a abajo a la izquierda y arriba a la derecha para el RP correspondiente a la primera epoch. En la Fig. 5.9 se puede apreciar la ampliación de estas zonas. Comparando con las figura del RP para la serie original del M-G (Fig. 5.2, se

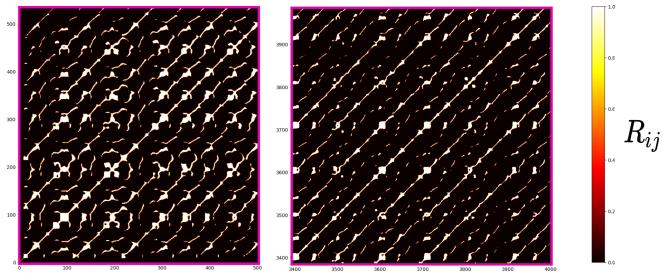


Fig. 5.9. Ampliaciones del RP para la *epoch 1* en el sistema M-G. **Izquierda:** cuadrado abajo izquierda. Todavía se puede leer bien un patrón similar al de M-G. **Derecha:** cuadrado arriba derecha. Al evolucionar nos encontramos con un patrón de diagonales paralelas indicando que la red está en una zona cercana el punto crítico.

puede ver como al principio del entrenamiento con plasticidad (Fig. 5.9 izquierda) se tiene un patrón parecido al correspondiente a $\rho_{sp} = 0.9$ ya que el radio espectral inicial se utiliza es $\rho_{sp} = 0.95$. Conforme va avanzando el sistema durante el primer *epoch* del entrenamiento, el patrón va cambiando debido a la regla de plasticidad dando lugar a líneas diagonales paralelas (Fig. 5.9 derecha). Este último RP es muy similar al que se ve en la Fig. 5.2 para $\rho_{sp} = 1.1$ pero, en este caso, el efecto esta puramente producido por la regla de plasticidad sináptica y la mejora que ésta introduce en la dinámica de la red.

6. CONCLUSIONES

Para el modelo básico de la ESN, se ha conseguido evidenciar la relación entre la dinámica interior de la ESN en torno al punto crítico y la mejora de la predicción como se ve en la Fig. 5.1. Esto también se ha comprobado con el análisis gráfico de los RP de la Fig. 5.2. Además, se ha calculado un umbral de criticalidad y unos parámetros óptimos de predictibilidad para los tres sistemas (Tab. 5.1) aunque nos hubiera gustado aumentar la estadística y afinar convenientemente la malla de parámetros para calcular los óptimos de forma más precisa.

Al aplicar la regla anti-Hebbiana, aunque solo ha sido posible obtener estadística para un entrenamiento de 10 *epochs* debido al elevado tiempo de computación, se han analizado varios fenómenos interesantes. Por un lado, en la Fig. 5.5 se ve como varían el radio espectral de la ESN, el MLE y el FPP en función del número de *epochs* mostrando que la regla de plasticidad hace que mejore la predicción en los primeros ciclos y se mantenga aproximadamente en la zona cercana al punto crítico (MLE ligeramente inferior a 0). Además, en la Fig. 5.6 se aprecia cómo mejora la dinámica de las trayectorias de las actividades neuronales en comparación a cuando no hay plasticidad ya que aumenta el recorrido de las neuronas por su espacio de fases. Al aumentar mucho

el entrenamiento con plasticidad (10 *epochs*, $\eta = 10^{-6}$), en la Fig. 5.7 se observa como la actividad neuronal se mueve por zonas diferentes y aproximadamente disjuntas. Esto indica que hay más de un atractor de la dinámica en la ESN y los estados neuronales incrementan su dependencia con estados anteriores del reservorio aumentando la inestabilidad de la ESN. Por último, usando de nuevo el análisis de recurrencia, se ha podido estudiar más en profundidad el efecto de la plasticidad en las activaciones gracias a los patrones que se presentan en las RP (Figs. 5.8 y 5.9). En estos RP se observa como al principio (pocos *epochs*) las activaciones se mueven en el espacio de fases en una cierta dirección sin repetir estados anteriores hasta que, con suficiente entrenamiento, la actividad neuronal se extiende por regiones amplias de dicho espacio (Fig. 5.8).

En cuanto a los objetivos del trabajo, se ha estudiado eficazmente la dinámica de la ESN y el efecto de la implementación de plasticidad sináptica como se acaba de comentar. Además, el uso del análisis recurrente ha permitido verificar los resultados y ha sido una herramienta útil para entender la dinámica de la ESN gráficamente. También se ha intentado estudiar el *representational drift* introduciendo un ruido gaussiano ξ en las activaciones según se discute en la Sec. 3.1. Sin embargo, no se han obtenidos resultados de valor para este estudio y por eso no se muestra nada sobre ello en el trabajo.

A título personal, nos gustaría hacer un par de comentarios sobre nuestra trayectoria realizando este proyecto. Teniendo en cuenta que partíamos con conocimientos aproximadamente nulos de python y Machine Learning, los primeros meses de trabajo los podríamos calificar suavemente como desmoralizantes. Las dificultades encontradas en la construcción de la base del código sí nos llevaron a nosotros a lo que podríamos llamar un punto crítico. Sin embargo, parece que interiorizamos este fenómeno de la ESN de mejor rendimiento al borde del caos y, justo a tiempo, solventamos estos errores de base y reflotamos nuestro trabajo. Una vez comenzado este proceso de investigación más exhaustivo y directo, empezamos a disfrutar y agradecemos haber escogido este trabajo de *Caos y ML*. Conforme avanzábamos íbamos descubriendo nuevos papers con enfoques distintos y nos intrigaba comprobar si podíamos reproducir alguno de esos resultados. Por esto nos tomamos la libertad de añadir el análisis de recurrencia que nos resultó peculiar y de gran interés. Nos hemos lamentado una y otra vez por los baches a los que nos enfrentamos en primer lugar porque, si hemos podido analizar todo esto en un intervalo de tiempo más bien reducido, nos hubiera gustado disponer algo más para continuar leyendo y aprendiendo sobre criticalidad dentro de la ESN. Sin embargo, entregamos este trabajo contentos y con ganas de seguir leyendo sobre estos y muchos más modelos de Sistemas Complejos en general y de ML en particular. Por último, agradecer a nuestro

tutor, Guillermo G. Barrios Morales, la atención, el material y la infinita paciencia.

REFERENCIAS

- [1] A Alonso, M De Curtis y R Llinás. «Postsynaptic Hebbian and non-Hebbian long-term potentiation of synaptic efficacy in the entorhinal cortex in slices and in the isolated adult guinea pig brain». En: *Proceedings of the National Academy of Sciences* 87.23 (1990), págs. 9280-9284.
- [2] Horace Barlow y Peter Földiàgk. «Adaptation and decorrelation in the cortex». En: ene. de 1989.
- [3] Giovanni Berlucchi y Henry A Buchtel. «Neuronal plasticity: historical roots and evolution of meaning». En: *Experimental brain research* 192 (2009), págs. 307-319.
- [4] Filippo Maria Bianchi, Lorenzo Livi y Cesare Alippi. «Investigating echo-state networks dynamics by means of recurrence analysis». En: *IEEE transactions on neural networks and learning systems* 29.2 (2016), págs. 427-439.
- [5] Joschka Boedecker et al. «Information processing in echo state networks at the edge of chaos». En: *Theory in Biosciences* 131 (2012), págs. 205-213.
- [6] Jean-Pierre Eckmann, S Oliffson Kamphorst, David Ruelle et al. «Recurrence plots of dynamical systems». En: *World Scientific Series on Nonlinear Science Series A* 16 (1995), págs. 441-446.
- [7] Benyamin Ghojogh y Mark Crowley. «The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial». En: *arXiv preprint arXiv:1905.12787* (2019).
- [8] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology press, 2005.
- [9] Hiroyuki K Kato, Ayako M Watabe y Toshiya Manabe. «Non-Hebbian synaptic plasticity induced by repetitive postsynaptic action potentials». En: *Journal of Neuroscience* 29.36 (2009), págs. 11153-11160.
- [10] Mantas Lukosevicius. «A practical guide to applying echo state networks». En: *Neural Networks: Tricks of the Trade: Second Edition* (2012), págs. 659-686.
- [11] Michael C Mackey y Leon Glass. «Oscillation and chaos in physiological control systems». En: *Science* 197.4300 (1977), págs. 287-289.
- [12] Norbert Marwan et al. «Recurrence plots for the analysis of complex systems». En: *Physics reports* 438.5-6 (2007), págs. 237-329.
- [13] Eduardo MAM Mendes y Erivelton G Nepomuceno. «A very simple method to calculate the (positive) largest Lyapunov exponent using interval extensions». En: *International Journal of Bifurcation and Chaos* 26.13 (2016), pág. 1650226.
- [14] Guillermo B Morales, Claudio R Mirasso y Miguel C Soriano. «Unveiling the role of plasticity rules in reservoir computing». En: *Neurocomputing* 461 (2021), págs. 705-715.
- [15] Guillermo B Morales y Miguel A Muñoz. «Optimal input representation in neural systems at the edge of chaos». En: *Biology* 10.8 (2021), pág. 702.
- [16] Michael E Rule, Timothy OLeary y Christopher D Harvey. «Causes and consequences of representational drift». En: *Current opinion in neurobiology* 58 (2019), págs. 141-147.
- [17] Steven H Strogatz. *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press, 2018.
- [18] Alan Wolf et al. «Determining Lyapunov exponents from a time series». En: *Physica D: nonlinear phenomena* 16.3 (1985), págs. 285-317.