



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Fenómenos a varias escalas en los  
mapeos autoorganizados y su relación con las  
capacidades de procesamiento de información

T E S I S

QUE PARA OBTENER EL TÍTULO DE:  
LICENCIADO EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A :

VÍCTOR MIRELES CHÁVEZ

TUTOR:  
M EN C. JOSÈ ANTONIO NEME CASTILLO



2008



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## AGRADECIMIENTOS

A todas las personas que en este tiempo me han mostrado lo bonito que puede ser el mundo.

A Toño Neme, por su amistad, asesoría, paciencia y apoyo, que han resultado invaluable para mi y sin quien nada de esto habría sido posible.

A Irma, Arturo, Lenia y Galia, mi familia, por su cariño, su apoyo y su comprensión.

A todos mis amigos, en especial a Adriana, Alejandra, Ana, Ariadna, Dalia, Daniela, Eric, Fátima, Hiroki, Irais, Irene, Ken, Lorena, Mariano, Maura, Miriam, Mónica, Natalia Carillo, Natalia Jonard, Natalia Martinez, Omar, Pozo, Rodrigo, Rosita, Rubén, Sonia y Yazmin, por su paciencia y comprensión.

A mi segunda familia: Abril, Paola, Memo y Daniel, por estar ahí siempre.

A mis sinodales, Guillermo, Omar, Luis y Raul, por sus valiosos comentarios y correcciones.

A mis profesores, Pedro Miramontes, Hector Mendez, Jeff King, Angel Tamariz, Angel Carrillo, y Jorge Urrutia, gracias a quienes encuentro en las matemáticas la mayor motivación.

A Aide, por compartir conmigo la felicidad sin reservas ni protocolos.

# Índice general

<b>1. Introducción</b>	<b>4</b>
<b>2. Preliminares</b>	<b>7</b>
2.1. Mapeos autoorganizados . . . . .	7
2.1.1. Motivación biológica . . . . .	8
2.1.2. Descripción detallada del algoritmo . . . . .	9
2.1.3. Medidas de error . . . . .	13
2.1.4. Aplicación al análisis de cúmulos . . . . .	14
2.2. Análisis de ondeletas . . . . .	15
2.2.1. El dominio de la frecuencia . . . . .	15
2.2.2. Transformada de Ondeletas . . . . .	17
2.2.3. Transformada discreta de Ondeletas . . . . .	20
<b>3. Fenómenos a varias escalas</b>	<b>22</b>
3.1. Ejemplos Espaciales . . . . .	25
3.2. Ejemplos en el SOM . . . . .	26
3.3. Aplicabilidad del análisis de ondeletas . . . . .	27
<b>4. Experimentos</b>	<b>29</b>
4.1. Disposición del experimento . . . . .	29
4.2. Conjuntos de datos usados en el SOM . . . . .	31
4.2.1. Uso de Codones (CODON) . . . . .	31
4.2.2. Ionósfera (ION) . . . . .	32

4.2.3.	Iris (IRIS) . . . . .	33
4.2.4.	Levadura (YEAST) . . . . .	33
4.3.	Series generadas con el SOM . . . . .	33
4.3.1.	Serie bidimensional . . . . .	34
4.3.2.	Series unidimensionales . . . . .	34
4.4.	Medidas Propuestas . . . . .	35
4.4.1.	Inverso del Rango de las correlaciones por nivel de las coordenadas X y Y de la serie bidimensional (CL2) . . . . .	35
4.4.2.	Inverso del rango de la media por nivel de los coeficientes para la serie unidimensional (IR) . . . . .	37
4.4.3.	Promedio por neurona de las varianzas (NMV) . . . . .	38
4.5.	Resultados . . . . .	38
<b>5.</b>	<b>Discusión General</b>	<b>47</b>

# Capítulo 1

## Introducción

Cuando leemos somos capaces de extraer el significado de una letra (por ejemplo una variable en una expresión matemática), de una palabra, de un enunciado, de un párrafo, o de un libro. Así pues, podemos decir que, tenemos funciones para mapear al conjunto de significados, el conjunto de letras, el de palabras, el de enunciados, el de párrafos y el de libros. El dominio de todas estas funciones es en realidad el mismo, sólo que visto con distinta resolución o granularidad. Por lo tanto un ente capaz de extraer significados de un texto debe poder leerlo con distintas resoluciones, justo aquellas que corresponden a las estructuras que componen el texto: letras, palabras, enunciados, etc. Si sólo pudiese leerlo, digamos, a la resolución correspondiente a letras individuales, jamás se daría cuenta de que cierto conjunto de ellas forma una palabra que tiene un significado asociado, distinto al de cada una de las letras.

Una forma de implementar esto en una computadora es leer el texto en la resolución más alta (letras), almacenarlo en memoria y después recorrer esta copia con distintos algoritmos, con resoluciones sucesivamente más bajas para reconocer estructuras de distintos tamaños. Un proceso de este tipo es necesario para la compilación y ejecución de código fuente escrito en un lenguaje de programación de alto nivel. Si bien en este caso todos los algoritmos leen carácter por carácter, conforme se desciende en resolución se utilizan localidades de memoria o estructuras de datos sucesivamente más grandes.

Los textos, sin embargo, no son los únicos objetos compuestos de estructuras de

distintos tamaños, o escalas. En el capítulo tres se dan ejemplos en varios contextos de objetos de este tipo. Una herramienta para reconocer estructuras de distintas escalas es el análisis de cúmulos, que consiste en el reconocimiento de subconjuntos de un conjunto de datos, de tal forma que los elementos de un subconjunto sean más parecidos entre sí, que respecto a los elementos de otros subconjuntos, en este caso la escala de un cúmulo es el número de elementos que lo componen.

Siguiendo esta misma línea, podemos hablar de fenómenos de varias escalas, tanto temporales como espaciales. Un buen ejemplo de esto es el juego de la vida, ideado Conway, en el que aparecen configuraciones de distintos períodos y que ocupan secciones de la rejilla de distintas áreas. Si observamos un sistema que tiene componentes moviéndose periódicamente con distintas frecuencias, podemos hablar de que dentro del sistema ocurren fenómenos de varias escalas temporales, que rigen el comportamiento de estos componentes. Análogamente si hay componentes que tienen rangos espaciales de movimiento de distintos tamaños decimos que se tienen fenómenos de varias escalas espaciales.

En este trabajo se estudia el comportamiento del mapeo autoorganizado (SOM), un modelo de la corteza cerebral con diversas aplicaciones, entre ellas el análisis de cúmulos. Este modelo se basa en ir modificando, sin supervisión, un mapeo  $\Lambda : \mathbb{R}^n \rightarrow \mathbb{R}^2$  de tal forma que, bajo las condiciones iniciales adecuadas, al final de la ejecución del algoritmo  $\Lambda$  envía grupos de elementos cercanos de nuestro conjunto de datos (contenido en  $\mathbb{R}^n$ ) a elementos cercanos de nuestro mapa (contenido en  $\mathbb{R}^2$ ). A esta propiedad se le conoce con el nombre de preservación de la topología. Las modificaciones al mapeo  $\Lambda$  se hacen sin supervisión alguna, en un proceso que por su naturaleza dinámica (discutida en la sección 2.1) se ha denominado de auto organización.

El comportamiento del SOM durante el proceso de auto organización no se ha comprendido del todo. En particular hasta ahora no existe prueba alguna de que el mapeo generado por el algoritmo en efecto converja a uno con las características deseadas[10]. Sería de interés cuantificar el comportamiento del algoritmo durante este proceso para intentar vislumbrar qué herramientas matemáticas son necesarias

para la prueba analítica de la convergencia a un mapeo preservador de la topología.

A lo largo de la ejecución del algoritmo podemos ir observando los cambios en el mapeo desde distintos puntos de vista, con el fin de saber las escalas espaciales y temporales de los fenómenos que ocurren. La tesis principal de este trabajo es que existe alguna relación entre las escalas de los fenómenos que ocurren dentro del SOM durante su ejecución y las escalas de los cúmulos que éste es capaz de reconocer.

Para medir lo primero, las escalas de los fenómenos dentro del SOM, se utiliza el análisis de ondeletas sobre series de tiempo generadas durante la ejecución. Para medir lo segundo, la capacidad del SOM de reconocer cúmulos de distintas escalas usamos las medidas estándar de error para el mapeo autoorganizado.

De esta forma se intenta indagar la relación entre las escalas de los fenómenos que ocurren dentro de este sistema procesador de información, y las escalas a las que es sensible en la información de entrada.

Las relación entre las escalas de los fenómenos que se presentan en un sistema procesador de información y sus capacidades de procesamiento de información se han estudiado previamente en otros sistemas, por ejemplo, se sugiere la relación entre el espectro de Fourier de los fenómenos en el juego de la vida y su propiedad de cómputo universal[39].



# Capítulo 2

## Preliminares

### 2.1. Mapeos autoorganizados

Los mapeos autoorganizados (SOM) son un tipo de red neuronal artificial de aprendizaje no-supervisado que fue presentado en 1982 por Kohonen[21]. Su funcionamiento se basa en una malla regular generalmente bidimensional de unidades o neuronas similares al perceptrón[38], en las cuales está distribuida la información acerca de un mapeo  $\Lambda : \mathcal{R}^n \rightarrow \mathcal{R}^2$ . Esta malla se va estimulando con vectores de entrada (estímulos) de un conjunto de características de dimensión  $n$ . Conforme esto ocurre la información contenida en las neuronas acerca del mapeo se va modificando por medio de dos procesos alternantes, uno de competencia y otro de cooperación, que da como resultado un mapeo  $\Lambda$  con algunas propiedades interesantes, de las cuales, la más destacada es su capacidad de preservar la topología del conjunto de estímulos, es decir, que estímulos cercanos se mapean en neuronas cercanas.

Es importante resaltar la naturaleza no-supervisada del SOM. A diferencia de otras redes neuronales artificiales, este algoritmo no compara constantemente su desempeño con algún desempeño ideal para ir corrigiéndose. Esto quiere decir que las modificaciones que va sufriendo el mapeo  $\Lambda$  no tienen que ver con que tan bueno sea; como ocurre, por ejemplo, con las redes neuronales conocidas como perceptrones multicapa[16] cuyo desempeño se compara constantemente con un desempeño ideal y en función de la discrepancia entre estos se va modificando la red.

El adjetivo autoorganizado del SOM es posiblemente su característica más sorprendente: al inicio de la ejecución del algoritmo, el mapeo  $\Lambda$  es aleatorio, y se va modificando sin supervisión alguna y sin coordinación central, hasta llegar a ser un mapeo capaz de preservar la topología. Las reglas que rigen la modificación del mapeo  $\Lambda$  son todas localmente definidas, es decir, en ningún momento las alteraciones se hacen en función del comportamiento global de  $\Lambda$ , sólo de su comportamiento al rededor de ciertos puntos. En pocas palabras la auto organización se define como el surgimiento de estructuras ordenadas sin supervisión global y por medio de una dinámica localmente definida. Ejemplos de este proceso se presentan en muchos campos [15], y los mapeos autoorganizados son un sencillo y eficaz modelo de uno de ellos: la formación de mapas en la corteza cerebral a partir de estímulos sensoriales.

### **2.1.1. Motivación biológica**

La experiencia de oír es el resultado de la estimulación de 24,000 células pilosas en cada oído, por acción de variaciones de presión en el aire. Estas células transforman las frecuencias detectadas por la membrana basilar en el oído en impulsos nerviosos que viajan hasta la corteza auditiva. Como se reporta en [44], sí se estimula el oído con una única frecuencia, cierta zona muy localizada de la corteza auditiva presenta actividad, y se puede ver que estas zonas están acomodadas en el campo auditivo cortical principal una tras otra de acuerdo a la frecuencia a la que son sensibles. Es decir, frecuencias parecidas se mapean en regiones contiguas de la corteza cerebral. Fenómenos semejantes se han reportado para las áreas corticales asociadas a la detección de fonemas, de rasgos ópticos elementales y a la localización geográfica [22].

En todos estos casos un estímulo consistente en la activación de miles o millones de terminales nerviosas se sigue de actividad en una región de la corteza. Es decir, puntos del conjunto multidimensional de estímulos se mapean en puntos del conjunto bidimensional de la corteza. Sin embargo el mapeo tiene la peculiaridad de que estímulos “cercanos” de acuerdo con algún criterio se mapean en puntos espacialmente cercanos en la corteza cerebral. Por mantener estas relaciones de distancia se dice que este mapeo preserva la topología. Incluso se han reportado respuestas de una sola

neurona a imágenes [25] o rostros humanos[46] determinados.

La idea de que estos mapeos, que preservan la topología, aparecen en el cerebro tras un proceso de auto organización puede rastrearse hasta los modelos computacionales sencillos de Willshaw y von der Malsburg[45], en donde se presentan ejemplos sencillos de redes biológicas que pueden ser formadas de esta manera. Kohonen[21] presenta un modelo conciso con sustento biológico de la formación de mapeos reductores de dimensión (cuantizadores de vectores) que preservan la topología en la corteza cerebral.

Si bien el SOM es efectivo en el sentido de que replica la preservación de la topología, éste ha recibido bastantes críticas. Por un lado se critica la preservación de la topología en sí, pues a simple vista parece contradictoria con la idea de la distribución de la información en el cerebro. Por otro lado, los patrones de activación en la corteza no son simétricos, mientras que en el algoritmo original del SOM sí lo son; en este sentido se han propuesto diversas modificaciones al SOM que no afectan su desempeño y su capacidad de preservar la topología pero que sí replican los patrones de actividad asimétricos de la corteza cerebral[27] [31]. Adicionalmente, casi desde su aparición se han propuesto modificaciones al algoritmo con la finalidad de mejorar su desempeño [20].

### 2.1.2. Descripción detallada del algoritmo

El algoritmo original descrito por Kohonen para generar un mapeo  $\Lambda : \mathfrak{R}^n \rightarrow \mathfrak{R}^2$  auto organizadamente se basa en una retícula bidimensional de  $N \times N$  unidades, o neuronas, cada una de las cuales almacena un vector  $v_{i,j} \in \mathfrak{R}^n$ , llamado vector de pesos. Llamemos  $\mathbb{G} = \{1 \dots N\} \times \{1 \dots N\}$  a la retícula de  $N \times N$  neuronas. Notemos que  $\mathbb{G} \subset \mathfrak{R}^2$ .

Con el vector de pesos que cada neurona tiene asociado se puede establecer una correspondencia entre algunos puntos en  $\mathfrak{R}^n$  (los vectores  $v_{i,j}$ ) y puntos en  $\mathbb{G} \subset \mathfrak{R}^2$  (las coordenadas  $i$  y  $j$ ). A lo largo del tiempo se va modificando esta correspondencia hasta que  $\Lambda$  mapee cierto conjunto  $S \subset \mathfrak{R}^n$  en  $\mathbb{G} \subset \mathfrak{R}^2$  preservando su topología, a  $S$  lo llamamos el conjunto de estímulos.

$\Lambda$  se define no solo para los puntos  $v_{i,j}$  ( $i, j$ )  $\in \mathbb{G}$ , se extiende a todos los puntos  $x \in \mathfrak{R}^n$  de la siguiente forma: dada  $x \in \mathfrak{R}^n$  definimos  $\Lambda(x) = a \in \mathbb{G}$  para la  $a$  que cumpla que  $\|v_a - x\| \leq \|v_b - x\| \quad \forall b \in \mathbb{G}$ .

Como queremos que ésta relación preserve la topología, es decir, que puntos cercanos en  $\mathfrak{R}^2$  estén asociados a puntos cercanos en  $\mathfrak{R}^n$ , a lo largo de la ejecución del algoritmo se deben modificar los vectores de pesos de forma que aquellos asociados a neuronas cercanas sean más o menos homogéneos. Una parte importante del SOM es la forma en que el concepto de "neuronas cercanas", o vecindades, varía con el tiempo. La capacidad de preservar la topología del mapeo va aumentando, refinándose con el tiempo. En [20] se divide la ejecución del SOM en dos etapas, la de ordenamiento y la de convergencia. En la primera etapa se organizará a grandes rasgos el mapeo para capturar las características gruesas del espacio de estímulos. En la segunda, se afinará el mapeo para captar a mayor detalle el espacio de estímulos. La división del algoritmo en estas etapas es, bajo las condiciones adecuadas, natural, es decir, no es necesario codificarlas explícitamente en el algoritmo ya que se presentan automáticamente durante la ejecución.

A lo largo de la ejecución de todo el algoritmo, descrita a continuación, se toman uno por uno los vectores del conjunto  $S$  y se selecciona el vector de pesos más cercano a él, el cual se modifica, junto con los vectores de pesos de las neuronas vecinas, acercándolo aún más. Recordemos que este concepto de vecindad va variando con el tiempo, y queda establecido por una función  $V : \mathbb{G} \times \mathbb{G} \times \mathbb{N} \rightarrow \mathfrak{R}$ , llamada función de vecindad que determina que tan cerca está una neurona de otra en un tiempo dado. Para  $a, b \in \mathbb{G}$  fijos,  $V(a, b, t_0) \geq V(a, b, t_1)$  si  $t_0 \leq t_1$ , es decir, es monótona no creciente en el tiempo. Como esta función mide la cercanía entre dos puntos en la retícula, es lógico pedir que  $\forall a \in \mathbb{G} \quad V(a, b, t_0) \leq V(a, a, t_0) \quad \forall b$ .

Definimos una época como un ciclo en el que tomamos todos los vectores de  $S$  una y sólo una vez para presentarlos al algoritmo y ajustar en base a ellos los pesos de nuestra malla. Llamemos  $T \in \mathbb{N}$  al número de épocas que dura la ejecución del algoritmo.

Al iniciar el algoritmo los vectores  $v_{i,j}$ , ( $i, j$ )  $\in \mathbb{G}$  se inicializan aleatoriamente.

Para cada  $t \in \{1 \dots T\}$  y para cada  $x \in S$  se encuentra un  $a_{x,t} = (i_{x,t}, j_{x,t}) \in \mathbb{G}$  tal que  $\|v_{a_{x,t}} - x\| \leq \|v_b - x\| \quad \forall b \in \mathbb{G}$ . Llamamos a esta  $a_{x,t}$  la BMU (best matching unit) o neurona ganadora. La norma que usamos es la norma  $L_1$  (norma de Manhattan). Una vez determinada la BMU, tomamos todas las neuronas y ajustamos su vector de pesos de acuerdo con:

$$v_b(t+1) = v_b(t) + V(a_{x,t}, b, t)\eta(t)(v_b(t) - x) \quad \forall b \in \mathbb{G} \quad (2.1)$$

donde  $\eta(t)$ , que llamamos función de aprendizaje, determina la magnitud de la alteración de los vectores de pesos de nuestra neurona conforme transcurre el tiempo.

En resumen: Dado  $N \in \mathbb{N}$  el tamaño del lado de la retícula,  $T \in \mathbb{N}$  el número de épocas que durará ejecutándose nuestro algoritmo y  $S \subset \mathbb{R}^n$  el conjunto de estímulos, se sigue el siguiente algoritmo:

1. Los vectores de pesos de las  $N \times N$  neuronas se inicializan aleatoriamente.
2. Para  $t$  desde 1 hasta  $T$

- a) Para cada  $x \in S$  encontrar  $a_{x,t} \in \mathbb{G}$  tal que  $\|v_{a_{x,t}} - x\| \leq \|v_b - x\| \quad \forall b \in \mathbb{G}$
- b) Modificar el vector de pesos de las neuronas de la siguiente forma:

$$v_b(t+1) = v_b(t) + V(a_{x,t}, b, t)\eta(t)(v_b(t) - x) \quad \forall b \in \mathbb{G}$$

El paso *a*), aquel en que se encuentra la BMU, lo llamamos de competencia pues podría decirse que todas las neuronas compiten para ver cuál tiene el vector de pesos más cercano al estímulo en cuestión. El segundo paso (*a*)) es de cooperación pues una neurona, al ser seleccionada, ocasiona que neuronas en su vecindad sean más parecidas al estímulo.

Dado este algoritmo podemos notar algunos parámetros ajustables: el tamaño  $N$  de la malla, la función de vecindad  $V$ , la función de aprendizaje  $\eta$  y el número de épocas  $T$ . La división entre etapa de ordenamiento y de convergencia se hace justamente en la evolución de las funciones  $\eta$  y  $V$ . Las funciones, abajo descritas, utilizadas en este trabajo cumplen con las características necesarias para que aparezcan éstas

dos etapas, a saber, que la función de aprendizaje no decrezca a cero pero que se mantenga por varias épocas en valores cercanos[16], esto se garantiza con su decrecimiento exponencial.

### Función de Vecindad

La función  $V : \mathbb{G} \times \mathbb{G} \times \mathbb{N} \rightarrow \mathfrak{R}$ , llamada función de vecindad determina qué tan cerca se encuentran a un tiempo  $t$  dos neuronas de la malla. Como se mencionó anteriormente, esta función es monótona no creciente en el tiempo[5], es decir, para  $a, b \in \mathbb{G}$  fijos,  $V(a, b, t_0) \geq V(a, b, t_1)$  si  $t_0 \leq t_1$  son dos tiempos distintos, y también decrece con la distancia euclidiana entre dos neuronas. Es decir, sean  $a = (i_a, j_a)$ ,  $b = (i_b, j_b)$  y  $c = (i_c, j_c) \in \mathbb{G}$  entonces  $\sqrt{(i_b - i_a)^2 + (j_b - j_a)^2} > \sqrt{(i_c - i_a)^2 + (j_c - j_a)^2} \Rightarrow V(a, b, t) < V(a, c, t)$ . Esta última restricción se ha relajado en distintas variantes del modelo, no necesariamente en detrimento del desempeño.

En este trabajo utilizamos una vecindad de las llamadas de tipo burbuja donde las neuronas sólo pueden estar dentro o fuera de la vecindad. Es decir, la función  $V$  sólo toma valores 0 o 1, dependiendo de si las neuronas están a una distancia mayor o menor a un cierto umbral  $U(t)$ . Este umbral va decreciendo con el tiempo según la función:

$$U(t) = De^{tk/T} \tag{2.2}$$

Donde  $D$  es el tamaño inicial de la vecindad y  $k = \log(E/D)$  con  $E$  el tamaño final.

### Función de Aprendizaje

La función  $\eta : \mathbb{N} \rightarrow \mathfrak{R}$  llamada función de aprendizaje determina el cambio en el tiempo del aprendizaje de las neuronas vecinas a la BMU. La idea es que esta función vaya decreciendo para ir haciendo cada vez más fino el reconocimiento de características del conjunto de estímulos. En este trabajo utilizamos una función de aprendizaje

$$\eta(t) = \eta e^{\frac{tq}{T}} \quad (2.3)$$

dónde  $\eta$  es el valor inicial de la función de aprendizaje, y  $q = \log(\frac{r}{\eta})$  con  $r$  el valor final.

De esta forma tenemos que hay tres constantes importantes:  $T, \eta, D$  que determinan, junto con los valores iniciales  $v_a$   $a \in \mathbb{G}$  que se inicializan aleatoriamente, cómo va a ser la ejecución de una instancia del SOM.

### 2.1.3. Medidas de error

Como queremos que los mapas generados con el algoritmo del SOM preserven la topología, podemos idear medidas de error que reflejen qué tan bien hacen esto. Una medida de error ampliamente utilizada en la literatura es el error de cuantización.  $EQ = \sum_{x_i \in S} (v_i - x_i)^2$  donde  $v_i = \Lambda(x_i)$  es el vector de pesos de la neurona asociada al estímulo  $x_i$ . Esto mide qué tanto se acercaron los vectores de pesos de las neuronas a las BMUs correspondientes.

Otra medida de error común que es la que usamos en este trabajo es el llamado error topográfico [18]

$$TE = \frac{1}{M} \sum_{x_i \in S} \delta(x_i, v_1, \dots, v_{n \times n}) \quad (2.4)$$

donde  $M = |S|$  y, si definimos la 2BMU para un estímulo como la segunda neurona con vector de pesos más parecida a él, tenemos:

$$\delta(x_i, v_1, \dots, v_{n \times n}) = \begin{cases} 1 & \text{si la BMU para } x_i \text{ no es adyacente a la 2BMU para } x_i \\ 0 & \text{en otro caso} \end{cases} \quad (2.5)$$

Esta última medida, a diferencia del error de cuantización, determina qué tanto la topología del espacio de estímulos se ve mapeada en la retícula  $\mathbb{G}$ , y es la medida que

se usa a lo largo de este trabajo, pues, como se discutirá más adelante, es también una buena medida de la capacidad del SOM para detectar cúmulos.

#### **2.1.4. Aplicación al análisis de cúmulos**

El análisis de cúmulos (Cluster Analysis) es una serie de métodos para intentar discernir, dentro de un gran conjunto de datos de alta dimensión, subconjuntos tales que los elementos de un subconjunto tengan algún rasgo en común. A grandes rasgos lo que se hace es definir una medida de la distancia entre los elementos, y se intenta encontrar una partición del conjunto en subconjuntos (cúmulos), tal que se minimice la distancia entre elementos de un mismo cúmulo y se maximice la distancia entre los centroides de cada cúmulo.

Si los datos son bidimensionales, y sólo en algunos casos, es fácil distinguir a simple vista los cúmulos. Pero cuando los datos son multidimensionales es imposible visualizarlos.

El análisis de cúmulos es una aplicación directa del SOM. Si un grupo de vectores en el espacio  $n$ -dimensional de estímulos está agrupado en un cúmulo, entonces tienden a mapearse todos en una misma neurona o en neuronas muy cercanas en la malla, preferiblemente adyacentes. En [24] se propone un método para auxiliar a un humano a interpretar un mapa generado por el SOM y detectar los cúmulos presentes en un conjunto de datos, este método consiste en colorear las neuronas de la malla según la proximidad de su vector de pesos con el de sus vecinas. Si éste está muy cerca de sus vecinas, se utiliza un tono oscuro. Si por el contrario está muy lejos de sus vecinas, se utiliza un tono claro. La escala de coloración que se utiliza es lineal. Los cúmulos, entonces, se observan en forma de regiones oscuras, rodeadas de regiones claras en la malla, asociadas estas últimas ya sea a estímulos aislados o a ningún estímulo.

El método de la coloración por escalas de grises hace muy fácil para un observador humano el reconocer cúmulos en el mapeo bidimensional generado por el SOM, sin embargo, para que éste tenga sentido, es indispensable la preservación de la topología ya que si ésta no se preserva, entonces los tonos de gris realmente no dicen nada sobre las relaciones de distancia existentes en el conjunto de estímulos.



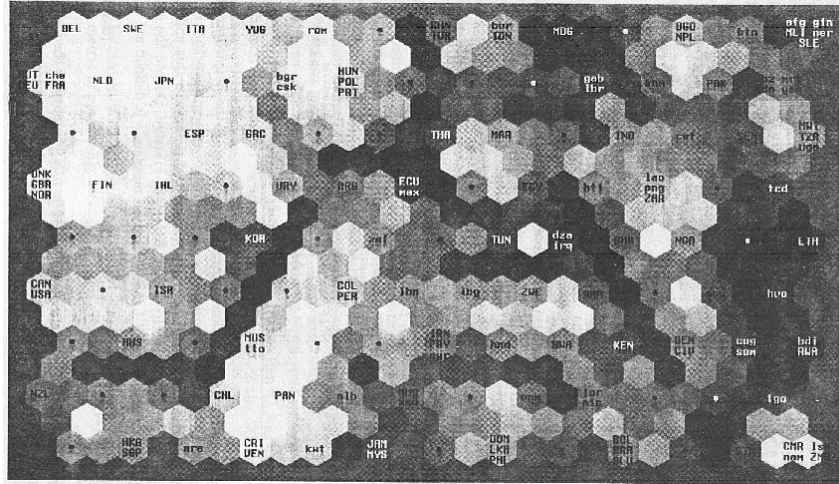


Figura 2-1: Mapa generado con el procedimiento descrito en [24] para el análisis de cúmulos usando el SOM. Imágen sacada de [20].

Ahora, para que el examinador humano pueda detectar cúmulos de distintos tamaños es necesario que se presenten en el mapa estas regiones claras rodeadas de oscuras en distintos tamaños y tonalidades de gris.

En este sentido es que el error topográfico descrito anteriormente es una medida adecuada de la capacidad para reconocer cúmulos.

## 2.2. Análisis de ondeletas

En este trabajo utilizaremos el análisis de ondeletas como una herramienta para analizar series de tiempo generadas durante el funcionamiento del SOM, con la finalidad de medir la diversidad de escalas en los comportamientos en ellas registradas. Para ello se da a continuación una explicación de los conceptos básicos de análisis de ondeletas.

### 2.2.1. El dominio de la frecuencia

El transformar una función cuyo dominio es el tiempo en una cuyo dominio es la frecuencia resulta útil para muchas aplicaciones[9]. Para entender a que nos referimos observemos la gráfica de  $f(x) = \sin(\frac{3}{5}x) + \sin(3x)$ .

Es interesante ver cómo una onda pareciera estar montada sobre otra onda de

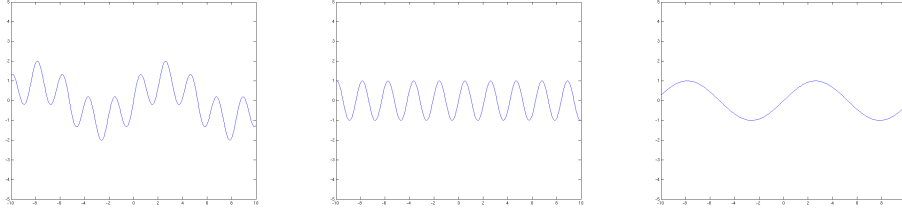


Figura 2-2: De izquierda a derecha, las gráficas de  $f(x) = \sin(\frac{3}{5}x) + \sin(3x)$ ,  $f(x) = \sin(\frac{3}{5}x)$ ,  $\sin(3x)$

menor frecuencia. En este caso  $f(x) = \sin(\frac{3}{5}x) + \sin(3x)$ , una función periódica, es, en efecto, la suma de dos funciones periódicas de distintas frecuencias. Una pregunta natural es que, si, dada una función periódica arbitraria de periodo  $a$ , es posible descomponerla como suma de funciones periódicas más sencillas, por ejemplo funciones seno y coseno. En efecto, basándose en el hecho de que la familia de funciones  $\{e^{inx} \mid n \in \mathbb{Z}\}$  forma una base ortonormal para las funciones cuadrado-integrables, la serie de Fourier nos permite, dada una función periódica  $f$  de periodo  $a > 0$  representarla por medio de un conjunto de coeficientes  $c(n)$   $n \in \mathbb{Z}$  y de la familia de funciones  $\{e^{inx} \mid n \in \mathbb{Z}\}$  por medio de la siguiente formula:

$$f(x) = \sum_{n \in \mathbb{Z}} c(n) e^{\frac{2\pi i n x}{a}} \quad (2.6)$$

Lo cual, utilizando la fórmula de Euler nos lleva a

$$f(x) = c(0) + \sum_{n \in \mathbb{Z}} (c(n) + c(-n)) \cos\left(\frac{2\pi n x}{a}\right) + \sum_{n \in \mathbb{Z}} i(c(n) - c(-n)) \sin\left(\frac{2\pi n x}{a}\right) \quad (2.7)$$

Esta descomposición transforma una función de  $\mathbb{R} \rightarrow \mathbb{C}$  cuyo dominio solemos interpretar como el tiempo, en un conjunto de coeficientes  $\{C(n) \mid n \in \mathbb{Z}\}$ , es decir, en una función  $\mathbb{Z} \rightarrow \mathbb{C}$  cuyo dominio puede interpretarse como la frecuencia, y cuyo contradominio está relacionado con la amplitud. En otras palabras, el conjunto  $\{C(n)\}$  de coeficientes de Fourier nos indica qué tanto la función  $f$  está compuesta por funciones sinusoidales de distintos periodos.

Esta descomposición es la base para el llamado análisis de Fourier, que eventual-

mente nos permite, por medio de la transformada discreta de Fourier analizar una serie de tiempo (en este trabajo consideramos a serie de tiempo como una función con dominio discreto) para determinar en qué grado ésta se asemeja al muestreo de una suma de funciones periódicas de distintos periodos. De esta forma, dada una serie de tiempo podemos obtener el llamado espectro de Fourier, una descomposición en el dominio de la frecuencia que nos dice la amplitud de las componentes de cada frecuencia de la serie de tiempo.

Una desventaja de la transformada discreta de Fourier es que los coeficientes que por medio de ella se calculan son globales. Es decir, si la serie de tiempo presenta un comportamiento de cierto periodo a lo largo de todo el dominio del tiempo, el coeficiente relacionado con este periodo es acertado, de lo contrario la fiabilidad de los coeficientes se reduce. De esta forma si tenemos una serie de tiempo que, digamos, en la primera mitad del dominio tiene un comportamiento de cierto periodo  $p_1$  y en la segunda mitad de otro periodo  $p_2$  la transformada discreta de Fourier intentará describir esta función como una complicada suma de funciones periódicas que reproduzca este comportamiento, sin tener en cuenta que la serie de tiempo es periódica por pedazos.

Una alternativa al análisis de Fourier es el análisis de ondeletas (Wavelet Analysis) [41][3]. Éste nos permite, dada una serie de tiempo, obtener una descomposición en el dominio de frecuencia cuya descripción es local. Es decir, es capaz de detectar qué tanto una serie de tiempo parece estar regida por una función de cierto rango de periodos en cierta región del dominio, independientemente del comportamiento en otras partes del dominio.

### 2.2.2. Transformada de Ondeletas

Recordemos [8] que en el espacio  $L^2$  de funciones reales cuadrado-integrables el producto interior entre dos funciones  $f$  y  $g$  está dado por:

$$\langle f, g \rangle = \int_X f(t)g(t)dt \quad (2.8)$$

donde  $X$  es el dominio de la función. En  $\mathbb{R}^n$  el producto escalar de dos vectores es proporcional a la norma de la proyección de uno sobre el otro. En este sentido, mide qué tanto dos vectores apuntan en la misma dirección.

Para hacer esto un poco más claro consideremos una familia ortonormal de funciones  $\mathcal{F} = \{\psi_a \mid a \in \mathbb{A}\}$  para  $\mathbb{A}$  un conjunto de índices. Recordemos que  $\mathcal{F}$  es ortonormal si

$$\forall a \in \mathbb{A} \quad \|\psi_a\|^2 = \langle \psi_a, \psi_a \rangle = \int_{\mathbb{R}} [\psi_a(t)]^2 dt = 1 \quad (2.9)$$

y

$$a, b \in \mathbb{A} \quad a \neq b \Rightarrow \langle \psi_a, \psi_b \rangle = 0 \quad (2.10)$$

Si tomamos ahora una función  $x(t) \in L^2$  que sea, por ejemplo, un múltiplo de una  $\psi_{a_0} \in \mathcal{F}$  i.e.  $x(t) = k\psi_{a_0}$  con  $k \in \mathbb{R}$ , entonces siendo (2.8) un producto interior y dada la condición de ortogonalidad (2.10) tenemos que  $\langle x, \psi_b \rangle = 0 \quad \forall b \in \mathbb{A} \quad b \neq a_0$  y  $\langle x, \psi_{a_0} \rangle = k$ , es decir, este producto interior nos ayuda a saber qué tanto  $x(t)$  es múltiplo de  $\psi_{a_0}$ , y dada la ortonormalidad de  $\mathcal{F}$  nos ayuda a aproximar, de ser posible, la función  $x(t)$  como una combinación lineal de elementos de  $\mathcal{F}$ . El requisito adicional para que esta descomposición siempre sea posible es que el espacio que  $\mathcal{F}$  genera,  $span(\mathcal{F}) = \left\{ \sum_{a \in \mathbb{A}} c_a \psi_a \mid c_a \in \mathbb{R} \right\}$ , sea denso en  $L^2$  en cuyo caso decimos que  $\mathcal{F}$  es una base de  $L^2$ .

Demos ahora a  $\mathcal{F}$  una forma que conviene a nuestros propósitos, tomemos al conjunto de índices  $\mathbb{A} = \mathbb{Z} \times \mathbb{Z}$ . Sea  $\psi \in L^2$  una función de soporte  $[-1/2, 1/2]$  y sean  $p, r \in \mathbb{Z} \quad r \neq 0$ . Definimos las funciones  $\psi_{r,p}(t)$  por medio de la siguiente expresión:

$$\psi_{r,p}(t) = \psi\left(\frac{t-p}{r}\right) \quad (2.11)$$

que es contracción/dilatación y translación de la función  $\psi$ . En esta expresión  $p$  especifica una translación y  $r$  una dilatación. Notemos que  $\forall r, p \in \mathbb{Z}$ ,  $\psi_{r,p}$  tiene como soporte el intervalo  $[(p-r)/2, (p+r)/2]$ . Asignemos al producto interior un nombre conveniente

$$CWT_x^\psi(r, p) = \langle \psi_{r,p}, x \rangle = \int_{\mathbb{R}} \psi_{r,p}(t)x(t)dt \quad (2.12)$$

El hecho de que el soporte de  $\psi_{r,p}(t)$  sea  $[(p-r)/2, (p+r)/2]$  nos permite interpretar  $CWT_x^\psi(r, p)$  como una propiedad local, es decir, nos habla de una característica que tiene nuestra función  $x(t)$  en un cierto intervalo, la característica de parecerse en cierto grado a la función  $\psi_{r,p}(t)$ . De esta forma, si mantenemos fijo a  $p$  y cambiamos  $r$ , el valor de  $CWT_x^\psi(r, p)$  nos dice qué tanto se parece nuestra función  $x(t)$  alrededor del punto  $p$  a distintas dilataciones de la función  $\psi$  centradas en  $p$ .

Si ahora cambiamos  $p$  entonces podemos saber qué tanto nuestra función  $x(t)$  se parece, en distintas localidades del dominio, a copias de distintos tamaños de la función  $\psi$ . Bajo ciertas propiedades (abajo descritas) de la función  $\psi$  lo anterior suele llamarse un análisis de tiempo y escala de la función  $x(t)$ , pues nos dice qué tanto se parece en cada tiempo nuestra función  $x(t)$  a la función  $\psi$  escalada de cierta forma. La escala se puede interpretar como el recíproco de la frecuencia, en el sentido de que un fenómeno de frecuencia alta se repite en escalas de tiempo pequeñas y viceversa.

El conjunto  $\{(\sin(2\pi nx), \cos(2\pi mx) \mid m, n \in \mathbb{Z})\}$  de funciones trigonométricas es también una base ortonormal del espacio de funciones cuadrado-integrables, con los coeficientes dados por la transformada de Fourier. Sin embargo, la base  $\mathcal{F}$  que hemos estado construyendo tiene ventajas sobre la base de funciones trigonométricas, a saber, que los coeficientes de la combinación lineal que representa a cada  $x \in L^2$  nos dan información local sobre su comportamiento en términos de escalas.

Recordemos que todos los elementos de la base  $\mathcal{F}$  son producto de translaciones y de dilataciones/contracciones de una función  $\psi \in L^2$ . Si esta función, además de tener soporte compacto, cumple con ser oscilatoria, es decir, que sea continua casi en todas partes y tenga valores tanto positivos como negativos, se le conoce como ondeleta (del francés ondelette, onda pequeña). Si además se le utiliza para construir una base ortonormal de  $L^2$  por medio de dilataciones/contracciones y translaciones la llamaremos la Ondeleta Madre, y las funciones  $\psi_{r,p}$  con  $r, p \in \mathbb{Z}$  Ondeletas Hijas.

Diversas ondeletas madre se han propuesto para generar bases ortonormales de

$L^2[3]$ . Entre las más usadas se encuentran las de Daubechies (en varias familias), Morlet, sombrero mexicano, y Haar.

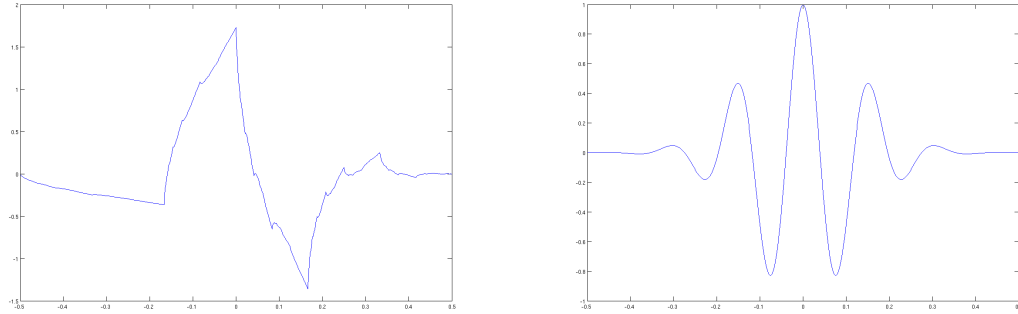


Figura 2-3: Ondeletas madre para las familias Daubechies 2 (Izquierda) y Morlet (Derecha)

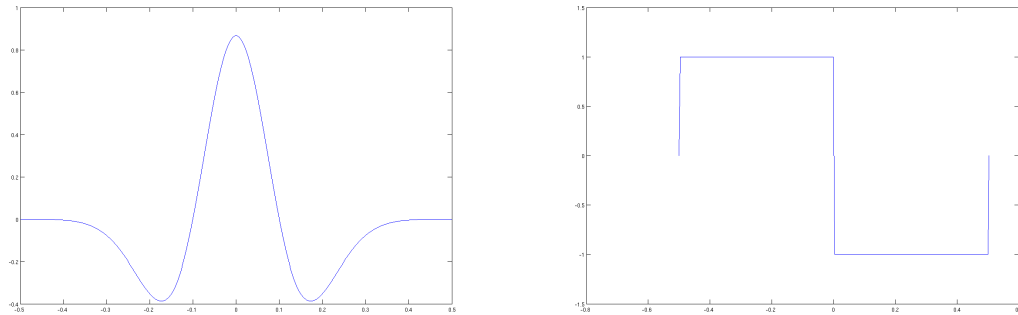


Figura 2-4: Ondeletas madre para las familias Sombrero Mexicano (Izquierda) y Haar (Derecha)

A la función  $CWT_x^\psi : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$  se le conoce como transformada continua de ondeletas de la función  $x$  correspondiente a la familia generada por la ondeleta madre  $\psi$ , y es una forma de análisis de tiempo y escala. A los valores  $CWT_x^\psi(r, p)$  se les conoce como coeficientes de ondeletas.

### 2.2.3. Transformada discreta de Ondeletas

El procedimiento arriba descrito nos dice cómo calcular los coeficientes de ondeletas para una función cuadrado-integrable. Así como la transformada discreta de Fourier es una aplicación de la transformada de Fourier que nos ayuda a analizar el

comportamiento en el dominio de la frecuencia de una serie de tiempo, la transformada discreta de ondeletas (DWT) nos ayuda a hacer un análisis de tiempo y escala de una serie de tiempo, o señal, es decir, de una función  $x(t) : \{1 \dots T\} \subset \mathbb{N} \rightarrow \mathbb{R}$ .

Notemos primero que 2.12 es semejante a una convolución, en el siguiente sentido:

$$CWT_x^\psi(r, p) = \int_{\mathbb{R}} \psi_{r,p}(t)x(t)dt = \int_{\mathbb{R}} \psi_{1,0}(p - \tau)x(\tau)d\tau = (x * \psi_{1,0})(p) \quad (2.13)$$

La transformada discreta de ondeletas se hace en términos de convoluciones discretas. Estas convoluciones se suelen interpretar como paso de la señal por un filtro, dónde la nueva función  $(x * \psi_{1,0})(p)$  se puede ver como la señal después de pasar por el filtro.

El algoritmo de cascada propuesto por Mallat[29] nos permite, pasando sucesivamente la señal por filtros de menor escala, calcular un conjunto de coeficientes de detalle y una constante, denominada coeficiente de aproximación. Los coeficientes de detalle están divididos en niveles, correspondientes a las escalas, de tal forma que los coeficiente de nivel 1 almacenan información sobre la escala más pequeña (fina) y los coeficientes de nivel  $n$ , información sobre la escala más grande (gruesa). El coeficiente de aproximación es el residuo entre esta aproximación finita y la función en sí.

Si la longitud de la serie de tiempo es  $2^n$ , la DWT nos devuelve para cada  $j \in \{1 \dots n\}$  el conjunto de coeficientes de detalle dado por

$$\{d_{j,i} \mid i \in \{1 \dots 2^{n-j}\}\} \quad (2.14)$$

Por supuesto no todas las series de tiempo tienen longitud igual a una potencia de dos. En este trabajo lo que hacemos es, dada una señal, agregarle ceros al final hasta completar una longitud igual a una potencia de dos.

# Capítulo 3

## Fenómenos a varias escalas

Para explicar a qué nos referimos con fenómenos a varias escalas podemos referirnos al juego de la vida de Conway, un autómata celular determinista bidimensional muy sencillo, en el que se han observado muchos fenómenos interesantes[11][17][34]. El conjunto de estados de este autómata tiene dos elementos, que se les denomina vivo y no vivo, llamémoslo  $\{1, 0\}$ . Si  $s$  es el estado actual y  $n$  el número de celdas en estado 1 en la vecindad de Moore (fig 3-1) de una celda, la función de transición que determina el siguiente estado de la celda es  $F : \{1, 0\} \times \{0 \dots 8\} \rightarrow \{1, 0\}$  definida como

$$F(s, n) = \begin{cases} 1 & \text{si } (s, n) \in \{(0, 3), (1, 2), (1, 3)\} \\ 0 & \text{de lo contrario} \end{cases} \quad (3.1)$$

Para esta discusión y para los ejemplos mostrados en este trabajo consideramos un juego de la vida con una retícula infinita. El espacio de configuraciones puede verse como puntos en  $\mathcal{X} = \{0, 1\}^{\mathbb{Z} \times \mathbb{Z}}$  el espacio de sucesiones bi-infinitas de sucesiones

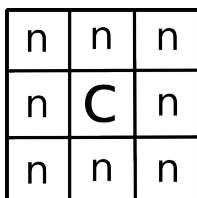


Figura 3-1: Las celdas marcadas con  $n$  son los integrantes de la vecindad de Moore de la celda marcada con  $c$



bi-infinitas de ceros y unos. Llamemos  $\Psi : \mathcal{X} \rightarrow \mathcal{X}$  al 9-código de bloques [28] generado por la función  $F$  descrita anteriormente, es decir, a la función que dada una configuración  $x_0 \in \mathcal{X}$ , nos da una nueva configuración  $\Psi(x_0) = x_1 \in \mathcal{X}$  los estados de cuyas celdas están dados por los estados de esa misma celda y sus vecinas en la configuración  $x_0$ , según la función  $F$ .

Aunque la discusión acerca de qué puede constituir un fenómeno en el juego de la vida puede tornarse filosófica, en este trabajo denominamos fenómeno cualquier sucesión de configuraciones del autómata  $\{x_i\}_{i \in \mathbb{N}}$  tal que  $x_{i+1} = \Psi(x_i)$ , es decir, la órbita de una configuración inicial  $x_0$ .

Para ejemplificar, antes de definir el concepto de escala espacial en el juego de la vida, podemos observar (fig. 3-2) algunas de las muchas configuraciones estables que se han encontrado; decimos que una configuración  $x \in \mathcal{X}$  es estable si es un punto fijo de  $\Psi$ , es decir  $\Psi(x) = x$ . Las configuraciones que se muestran y las subsecuentes son tomadas de [43], un excelente catálogo de fenómenos interesantes en el juego de la vida.

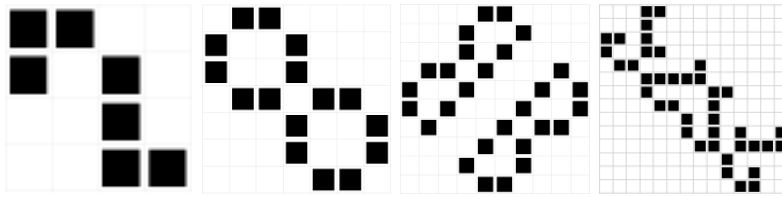


Figura 3-2: Configuraciones estables de (de izquierda a derecha) 7, 16, 28 y 40 celdas

Como se observa hay configuraciones estables de varios tamaños, es decir, para generar cada uno de los que se muestra es necesario especificar los valores de un número distinto de celdas. Una vez especificado el valor de estas celdas la configuración en ellas es invariante bajo la función  $\Psi$ .

Así como existen configuraciones estables en el juego de la vida, hay también configuraciones periódicas de muchos periodos. Decimos que una configuración  $x$  es de periodo  $p$  si  $p$  es el número natural más pequeño tal que  $\Psi^p(x) = x$  donde para  $k \in \mathbb{N}$  definimos

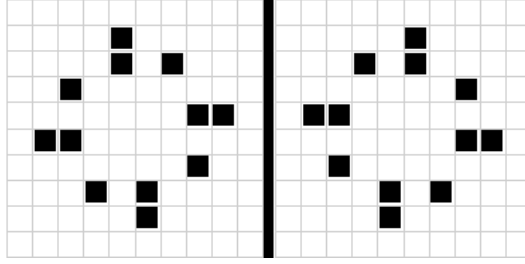


Figura 3-3: Un oscilador de periodo 2 cuyas dos configuraciones constan de 12 celdas vivas y está contenido en una región de  $10 \times 10$  celdas

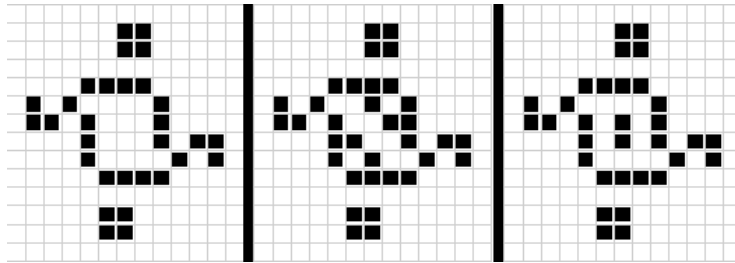


Figura 3-4: Un oscilador de periodo 3 cuyas configuraciones constan de 30, 34 y 32 celdas vivas y está contenido en una región de  $13 \times 14$  celdas

$$\Psi^k(x) = \begin{cases} x & \text{si } x = 0 \\ \Psi(\Psi^{k-1}(x)) & \text{si } x > 0 \end{cases} \quad (3.2)$$

A las configuraciones periódicas también se les conoce como osciladores. En la literatura se han reportado osciladores de periodos 2, 3, 4, 5, 6, 7, 8, 9, 12, 16, 14, 16, 22, 30, 32, 46, 54, 60, 100 y 144. En las figuras 3-3, 3-4 y 3-5 se muestran tres ejemplos de osciladores de distintos periodos.

Dado el fenómeno  $\{x_0, x_1, x_2, \dots\}$  con una cierta configuración inicial  $x_0 \in \mathcal{X}$ ,

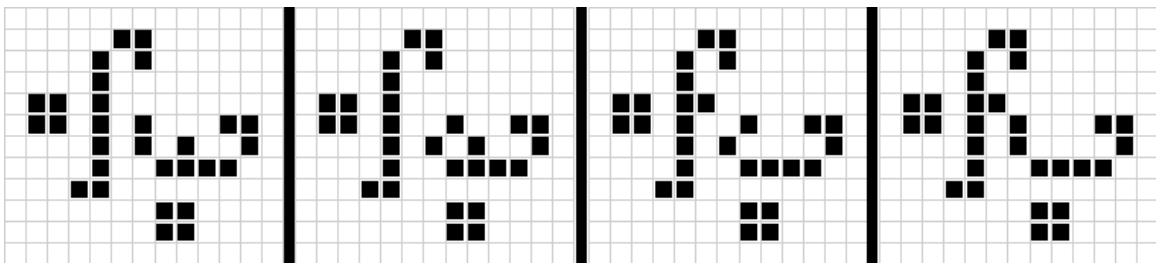


Figura 3-5: Un oscilador de periodo 4, para el cual todas las configuraciones constan de 29 celdas vivas y que está contenido en una región de  $12 \times 13$  celdas

podemos contar el número de 1's en cada  $x_i$  y el máximo de estos números lo denominamos la escala espacial de este fenómeno. Es claro que podrían existir configuraciones cuya escala sea infinita, aquellas que tienden a abarcar una región no acotada de la malla. Por otro lado la escala temporal la definimos como el número de configuraciones distintas en la órbita en cuestión. De esta forma, una configuración estable tiene una escala temporal uno, mientras que un oscilador de periodo  $p$  tiene una escala temporal  $p$ .

Este sistema de espacio y tiempos discretos sirve para ilustrar los conceptos de escalas espaciales y temporales, sin embargo, estas características se pueden asignar a fenómenos en muchos otros sistemas.

### 3.1. Ejemplos Espaciales

Un buen ejemplo de un sistema que presenta fenómenos de distintas escalas espaciales es el que se menciona en la introducción de este trabajo, aquel que se refiere a las distintas resoluciones con que podemos leer una cadena de símbolos para extraerle significado. Esto habla de que la organización del texto tiene distintas escalas espaciales, desde la unitaria (letras) hasta escalas muy grandes como los párrafos. Sin embargo se pueden encontrar muchos otros ejemplos, entre ellos

1. Los tamaños de los pedazos de hielo marino que se desprenden de los polos son varios[23] y el tamaño de cada bloque de hielo determina la forma en que éste presentará grietas causadas por colisiones con otros bloques[26].
2. La diferenciación genotípica que resulta de la selección positiva de la tolerancia a metales pesados en poblaciones de plantas se reporta en escalas desde algunos centímetros hasta varios kilómetros[42].
3. Los patrones de forrajeo y en general la distribución de animales en busca de recursos se presentan también en varias escalas, es decir, hay saltos de varios tamaños. Más aún, se reporta que la distribución de frecuencias de los distintos tamaños de saltos es una distribución de Levy[2][37][35].

4. La conformación tridimensional de las proteínas suele dividirse en 4 niveles de descripción. Los elementos de la descripción de la estructura de cada nivel son sucesivamente más grandes, desde aminoácidos hasta cadenas polipeptídicas de varios miles de aminoácidos de longitud. La dinámica, aún por comprenderse, que determina la estructura tridimensional de una proteína parece estar definida en términos de elementos de distintos tamaños.

## 3.2. Ejemplos en el SOM

En el algoritmo tradicional del SOM, los tamaños de las regiones que se ven afectadas por cada estímulo están explícitamente codificadas en la función de vecindad, y los tamaños de estas regiones en efecto varían. Algunos resultados sugieren que lo que necesita decrecer en la vecindad es el número de neuronas en ella y no el radio[32]. En todo caso, el tamaño, ya sea medido en número de elementos o en radio, de las regiones que se activan varía con el tiempo. De la misma forma con el parámetro de aprendizaje se hace variar explícitamente el tamaño de las alteraciones de los vectores de peso.

Pero estos no son los únicos fenómenos que presentan ese comportamiento. En el ámbito espacial, como se verá más adelante (sección 4.3.2), la diferencia entre el vector de pesos de la BMU y el estímulo que la activa, también es de distintos tamaños, esto es independiente de cualquiera de los parámetros explícitamente decrecientes.

En el ámbito temporal se observó que en una ejecución del SOM, la distancia entre los vectores de pesos de algunas neuronas y los estímulos que a ellas se mapean presenta un comportamiento aparentemente oscilatorio de un periodo distinto al número de estímulos. Si uno presenta los estímulos al SOM siempre en el mismo orden también se observaría una periodicidad igual al número de estímulos, principalmente hacia el final de la ejecución.

### 3.3. Aplicabilidad del análisis de ondeletas

Cuando decimos que un cierto fenómeno tiene una cierta escala temporal intuitivamente significa que su duración es de cierto orden de unidades de tiempo, lo cual se puede ver como que el soporte de la función que lo describe es de cierto tamaño. Si esto es así, la función es más parecida a una función cuyo soporte es de ese mismo orden que a alguna cuyo soporte sea más pequeño o más grande. Esta comparación con funciones de soportes de distintos tamaños se lleva a cabo en el análisis de ondeletas. Adicionalmente el análisis de ondeletas es local en el tiempo, es decir, esta información varía conforme analizamos distintas regiones del dominio del tiempo.

Por lo anterior el análisis de ondeletas es un análisis de tiempo-frecuencia, sin embargo el término *frecuencia* raramente se utiliza en este contexto. En su lugar se utiliza el concepto de *escala*, pues siendo las ondeletas funciones no periódicas lo que los coeficientes en realidad miden es qué tanto se parece la función (o serie de tiempo), en distintas regiones del dominio, a la ondeleta madre escalada de cierta manera, es decir, con una cierta longitud de soporte.

Observar las diferencias de los coeficientes de ondeletas de distintas escalas (niveles, en la terminología de la transformada discreta de ondeletas), nos da una visión sobre las escalas temporales de las series de tiempo. Por otro lado, la información sobre las escalas espaciales está contenida en las magnitudes de todos los coeficientes sin importar el nivel, pues nos dicen el tamaño en el dominio del espacio, de las distintas componentes de la serie de tiempo.

Es importante resaltar el carácter local de la transformada de ondeletas, porque puede detectar una variación de cierta escala en un segmento de la serie de tiempo sin importar que a lo largo de la serie se presenten variaciones diferentes en esa escala. Es esta la razón por la que, en este trabajo, se utilizó el análisis de ondeletas y no el análisis de Fourier.

En este trabajo se propone el uso del análisis de ondeletas para el estudio del comportamiento del SOM. Esto se hará por medio del análisis de los coeficientes de ondeletas de series de tiempo generadas durante la ejecución del SOM, con la finalidad

de determinar la diversidad de las escalas espaciales y temporales de los fenómenos registrados en estas series.

# Capítulo 4

## Experimentos

El objetivo de este trabajo es analizar el comportamiento del mapeo autoorganizado (SOM) durante el proceso de auto organización. Para ello se ejecutó el algoritmo del SOM 4,000 veces sobre cada uno de 4 conjuntos de datos y se registraron, en un conjunto de series de tiempo, algunos de los fenómenos que se presentaron en las variables de estado del SOM durante cada ejecución. Estas series de tiempo fueron posteriormente analizadas usando la transformada discreta de ondeletas para determinar las escalas espaciales y temporales ahí registradas. La variedad de estas escalas se midió por medio de tres medidas aquí propuestas, y estas fueron comparadas con la medida de error topográfico (TE) para el mapeo correspondiente, para intentar inferir alguna relación entre la variedad de escalas en los fenómenos dentro del SOM y el desempeño del algoritmo. Este procedimiento se esquematiza en la figura 4-1.

### 4.1. Disposición del experimento

En este trabajo se utilizaron cuatro conjuntos de datos, comunmente utilizados para realizar pruebas a algoritmos de análisis de cúmulos, éstos son descritos en la siguiente sección. Sobre cada uno de estos conjuntos de datos se ejecutó el algoritmo del SOM, descrito en la sección 2.1.2, 4000 veces, 2000 de ellas utilizando siempre el mismo orden para presentar los estímulos, y las otras 2000 presentando los estímulos en orden aleatorio de forma que en cada época se presenten todos los estímulos una

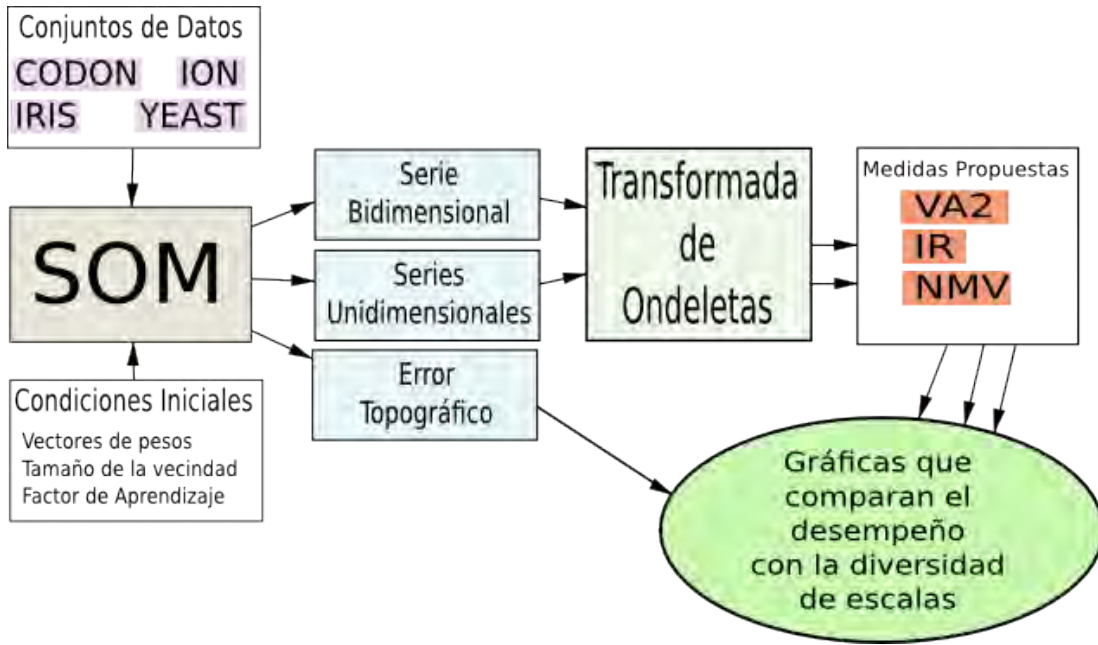


Figura 4-1: Esquema general de los experimentos

y sólo una vez. Todas las ejecuciones tienen retículas de  $20 \times 20$  (es decir  $|\mathbb{G}| = 400$ ) y vecindades de tipo burbuja. La norma utilizada en  $\mathbb{G}$  es la norma  $L_1$  o norma de Manhattan. La norma utilizada en  $(R)^n$  es la norma euclidiana. El objetivo de realizar tantas ejecuciones sobre cada conjunto de datos es para registrar el comportamiento del algoritmo con distintas condiciones iniciales. Para cada ejecución tenemos una serie de parámetros, que son escogidos al azar para cada ejecución:

1.  $T$  número de épocas.
2.  $U_{inic}$  El umbral inicial de la burbuja de la vecindad.
3.  $\eta_{inic}$  El valor inicial del parámetro de aprendizaje.
4. El valor inicial de los vectores de pesos de las neuronas.

El valor final del umbral de la vecindad y del parámetro de aprendizaje siempre es 0,00001. El umbral de la vecindad decrece según la función

$$U(t) = U_{inic} e^{tk/T} \quad (4.1)$$



donde  $k = \log\left(\frac{0,00001}{U_{inic}}\right)$ , de esta manera la función  $U(t)$  describe un decrecimiento exponencial desde el valor  $U_{inic}$  hasta 0,00001. Análogamente el valor del factor de aprendizaje decrece con la función

$$\eta(t) = \eta_{inic} e^{tq/T} \quad (4.2)$$

con  $q = \log\left(\frac{0,00001}{\eta_{inic}}\right)$

Para cada una de estas ejecuciones se calcula el error topográfico (TE), y se obtiene un conjunto de series de tiempo, a las cuales, como se describe en la sección 4.4, se les somete a la transformada discreta de ondeletas para finalmente calcular, para cada ejecución, una serie de medidas de la diversidad de las escalas de los fenómenos en ella presentados.

Cada ejecución del algoritmo podría tomar más épocas de las necesarias, es decir, el mapeo puede converger a una configuración estable de dónde ya no salga por el resto de las épocas. Puesto que en este trabajo lo que se desea es conocer el comportamiento del SOM durante el proceso de auto organización, es decir, justamente en las épocas en que el mapeo transita hacia el estado estable, las series de tiempo generadas se truncan en el tiempo en que el mapeo entra en una configuración estable (si es que esto sucede antes de transcurridas  $T$  épocas).

## 4.2. Conjuntos de datos usados en el SOM

### 4.2.1. Uso de Codones (CODON)

El ADN codificante contiene información sobre las secuencias de aminoácidos que conformarán las proteínas del organismo en cuestión. Esta información está contenida en forma de palabras de tres pares de bases de longitud, denominadas codones. De esta forma, una secuencia de ADN codificante se puede ver como una secuencia de codones que será traducida en una secuencia de aminoácidos que conformarán las proteínas. Antes de que esta traducción pueda llevarse a cabo, el ADN es transcrito en ARN el cual llega al ribosoma, que es dónde se realiza la traducción. En este proceso, a

cada codón se adhiere un tipo específico de molécula de ARN de transferencia que está enlazada a un aminoácido el cual a su vez se enlaza con los otros aminoácidos que se han traducido previamente. Lo que determinan qué tipo de RNA de transferencia, y por lo tanto qué tipo de aminoácido, es representado por cada codón es lo que se conoce como código genético, un diccionario que sirve, eventualmente, para traducir una secuencia de ADN a proteínas

Por su relación con la síntesis de proteínas, el estudio de la frecuencia de uso de cada uno de los 64 codones en el genoma de un organismo, se ha intentado utilizar para determinar qué secciones del ADN son codificantes y cuales no [6], y para distinguir genes altamente expresados [13]. Sin embargo un genoma completo, no únicamente la parte codificante, puede ser visto como una larga cadena de codones, y el estudio de estas frecuencias se ha propuesto como un método para clasificar genomas [12].

Este conjunto de datos consta de la frecuencia de cada uno de los 64 codones en los genomas de 102 organismos. Es por lo tanto un conjunto de 102 puntos contenidos en  $\mathbb{R}^{64}$ . Como se reporta en [33] existe cumularidad en este conjunto de datos, la cual está relacionada con la filogenia de los organismos.

#### 4.2.2. Ionósfera (ION)

La ionósfera normalmente refleja las ondas de radio, y este hecho suele usarse para la transmisión de las llamadas señales de onda corta. Adicionalmente, esto permite el uso de radares para investigar la estructura de la ionosfera. Sin embargo, no todas las emisiones de los radares son satisfactoriamente reflejadas por la ionosfera. Este conjunto de datos proviene de mediciones simultáneas de 17 radares en Labrador, Canadá emitiendo con una potencia de 6.4 KW, algunas de las 350 emisiones realizadas son insatisfactorias. Cada lectura de cada radar es cuantificada con dos números correspondientes al valor de una función compleja que toma como argumento las mediciones del radar. Esto nos da un total de 34 números por emisión, para un total de 350 emisiones[40]. Algunas de ellas se clasifican como buenas y otras como malas, aunque ésta última característica no fue alimentada a nuestro algoritmo. Por lo anterior este conjunto de datos es de longitud 150 y está contenido en  $\mathbb{R}^{34}$ .

### 4.2.3. Iris (IRIS)

Existen 3 especies de la planta llamada Iris: *Iris setosa*, *Iris virginica* e *Iris versicolor*. Este conjunto de datos, recopilado por Anderson en 1935, toma mediciones de 50 especímenes de cada una de ellas. Las mediciones que se toman son el ancho del pétalo, el largo del pétalo, el ancho del sépalo, y el largo del sépalo, lo que hace de este un conjunto uno de 150 vectores en  $\mathbb{R}^4$ . [7]

### 4.2.4. Levadura (YEAST)

La levadura de cerveza (*Saccharomyces cerevisiae*) tiene dos procesos metabólicos interesantes. El primero, cuando se le sitúa en un medio rico en glucosa, es el de fermentación, en el que transforma la glucosa en etanol. En el segundo, una vez que la glucosa es insuficiente, la levadura cambia a la respiración aerobia. Este cambio de metabolismo es resultado de un cambio en la expresión de los genes de la levadura[4]. Este conjunto de datos contiene la expresión de los 7 genes involucrados en 834 momentos distintos. Se trata entonces de un conjunto de 834 puntos en  $\mathbb{R}^7$

## 4.3. Series generadas con el SOM

Las series descritas a continuación se generan a lo largo de la ejecución del SOM. Es importante recordar que en la mayoría de los casos el mapeo que se va generando con el SOM converge antes de que la ejecución del algoritmo termine. Para que esto no haga que el análisis de ondeletas detecte un comportamiento periódico hacia la cola de la serie, se cortó la serie a partir de la primera época en que el mapeo no se alteró, es decir, en que los estímulos fueron mapeados a las mismas BMUs que en la época anterior. Una vez hecho esto, se ejecuta la transformada discreta de ondeletas (bidimensional o unidimensional según sea el caso) sobre la serie resultante y se calculan diversas medidas de variación de escala sobre los coeficientes de detalle resultantes. Las medidas propuestas se discuten en la siguiente sección.

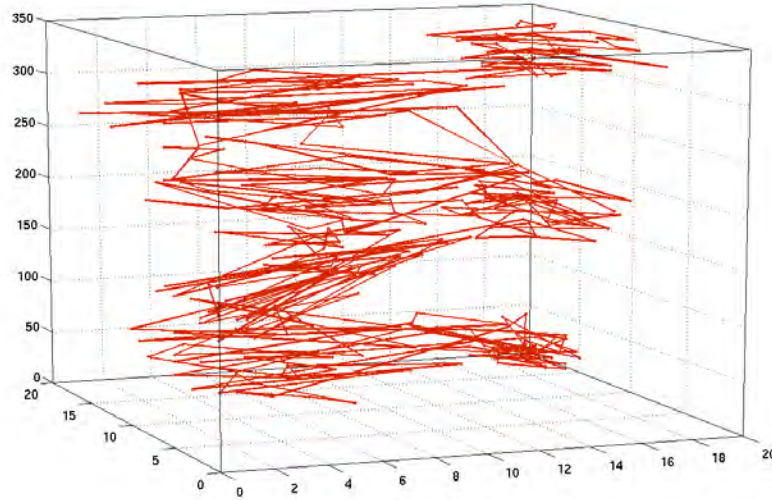


Figura 4-2: Serie bidimensional para una ejecución sobre el conjunto Iris. La retícula  $\mathbb{G}$  está en el plano XY, el tiempo en el eje Z.

### 4.3.1. Serie bidimensional

Esta serie registra el movimiento de la BMU a lo largo del tiempo, sin importar para qué estímulo se haya seleccionado. Si  $T$  es el número total de épocas y  $M$  el número de estímulos (longitud del conjunto de datos), esta serie tiene longitud  $TM$ . Es bidimensional pues cada elemento de la serie consta de las dos coordenadas de la neurona correspondiente.

Esta serie de tiempo puede verse como una función  $\mathcal{C} : \{1 \dots TM\} \rightarrow \mathbb{N} \times \mathbb{N}$  definida como

$$\mathcal{C}(t) = (x_t, y_t) \tag{4.3}$$

donde  $(x_t, y_t)$  son las coordenadas de la BMU al tiempo  $t$ .

Un ejemplo puede verse en la figura 4-2.

### 4.3.2. Series unidimensionales

Por cada una de las 400 neuronas, podemos ver en que momentos ésta fué seleccionada como BMU. Cada vez que esto ocurre su vector de pesos se altera según (2.1). Esta alteración depende del parámetro de aprendizaje  $\eta$  que varía con el tiempo, y

también de la diferencia entre el vector de pesos de la neurona actual y el estímulo para el que fue seleccionada como BMU. Esta diferencia tiende a decrecer conforme el mapeo se acerca a su estado final estable, por lo que podría verse como una medida de qué tan lejos de su estado final se encuentra el vector de pesos de la neurona. Eliminamos de esta medición el parámetro de aprendizaje para no registrar en la serie de tiempo el decrecimiento artificial que induce. Si bien esta diferencia tiende a disminuir conforme el mapeo converge, esta tendencia puede verse afectada si la misma neurona es seleccionada como BMU por dos estímulos distintos, lo cual ocurre muy a menudo.

Para cada  $a \in \mathbb{G}$  definamos  $T_a \subset \mathbb{N}$  como el conjunto de instantes de tiempo en que  $a$  fue seleccionada como BMU. Definamos la función de ordenación  $S_a : \{1 \dots |T_a|\} \rightarrow T_a$  tal que  $S_a(t) < S_a(s) \Leftrightarrow t < s$ . Definimos entonces la serie unidimensional como

$$\mathcal{D}_a(t) = d(x(S_a(t)), v_a(S_a(t))) \quad t \in \{1 \dots |T_a|\} \quad (4.4)$$

donde  $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  es la distancia euclidiana,  $x(s)$  es el estímulo al tiempo  $s$  y  $v_a(s)$  es el vector de pesos de la neurona  $a$  al tiempo  $s$ . En general  $|T_a|$  es distinto para cada  $a \in \mathbb{G}$ , y por lo tanto el máximo nivel de descomposición de ondeletas que cada una de estas series acepta es distinto. Cuatro ejemplos de este tipo de serie se muestran en la figura 4-3.

## 4.4. Medidas Propuestas

### 4.4.1. Inverso del Rango de las correlaciones por nivel de las coordenadas X y Y de la serie bidimensional (CL2)

Con la finalidad de detectar fenómenos bidimensionales de distintas escalas, tomamos la serie bidimensional y medimos las correlaciones entre los coeficientes de ondeletas para sus coordenadas X y Y. Proponemos el cálculo de esta correlación como un método para detectar cuando está ocurriendo un fenómeno bidimensional. De esta forma, una correlación alta a un cierto nivel nos dice que existen fenómenos bidimensionales

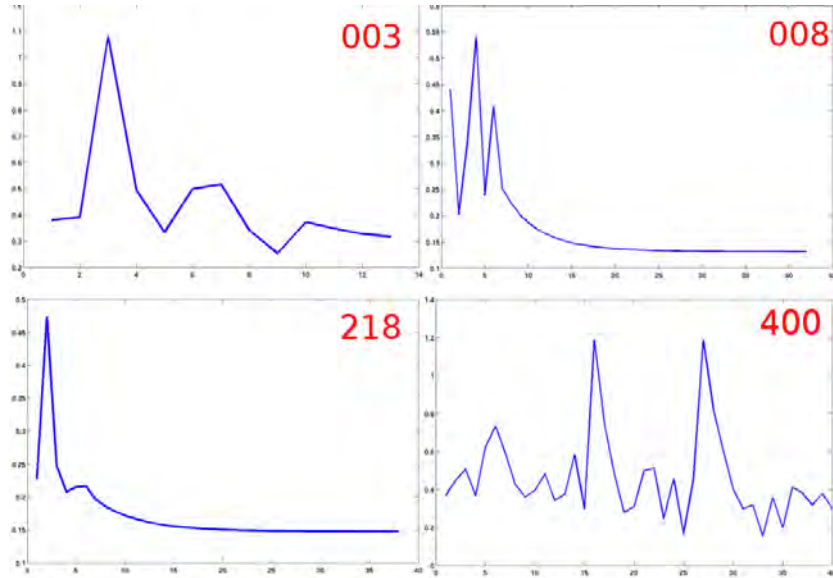


Figura 4-3: Para las neuronas 3, 8, 218 y 400 sobre una ejecución sobre el conjunto Iris, en el eje horizontal el tiempo, en el vertical la diferencia entre el vector de pesos y el estímulo que la activa como BMU.

de la escala temporal correspondiente. Puesto que no nos interesa si la correlación es negativa o positiva, tomamos el valor absoluto de los índices de correlación de Pearson[30].

La longitud del intervalo en el que estas correlaciones por nivel están contenidas, nos proporciona una medida de que tan alejadas están, es decir, si hay una gran disparidad entre el más grande y el más pequeño. Como las correlaciones son números entre 0 y 1 tenemos que la longitud de este intervalo es a lo más 1. Si ahora nos fijamos en el recíproco de esta longitud, tenemos una medida de qué tan juntas están las correlaciones de cada nivel, es decir, de qué tan homogénea es la distribución de los fenómenos bidimensionales a lo largo de las diferentes escalas temporales.

Dada una serie de tiempo  $\mathcal{C}(t) = (x_t, y_t) \ t \in \{1 \dots T\}$ , lo primero que hacemos es calcular por medio de la transformada discreta de ondeletas, los coeficientes de detalles de todos los niveles para sus coordenadas X y Y, lo que nos da como resultado dos conjuntos  $D_x = \{d_{j,i}^x \mid i \in \{1 \dots 2^{n-j}\}\}$  y  $D_y = \{d_{j,i}^y \mid i \in \{1 \dots 2^{n-j}\}\}$  con  $j \in \{1 \dots n\}$  donde  $n = \lceil \log_2(T) \rceil$ . Enseguida, para cada  $j \in \{1 \dots n\}$  calculamos el coeficiente de correlación:

$$CO_j = \frac{2^{n-j} \sum d_{j,i}^x d_{j,i}^y - \sum d_{j,i}^x \sum d_{j,i}^y}{\sqrt{2^{n-j} \sum (d_{j,i}^x)^2 - (\sum d_{j,i}^x)^2} \sqrt{2^{n-j} \sum (d_{j,i}^y)^2 - (\sum d_{j,i}^y)^2}} \quad (4.5)$$

donde las sumas son para  $i$  desde 1 hasta  $2^{n-j}$ . Este número es conocido como el coeficiente de correlación de Pearson y mide qué tanto el cambio entre dos variables está linealmente correlacionado, en el sentido de qué tan común es que cuando una de ellas cambie la otra también lo haga. Una vez obtenidos estos coeficientes, definimos la medida CL2 de la siguiente forma

$$CL2 = \frac{1}{\max_{j \in \{1 \dots n\}} \{|CO_j|\} - \min_{j \in \{1 \dots n\}} \{|CO_j|\}} \quad (4.6)$$

#### 4.4.2. Inverso del rango de la media por nivel de los coeficientes para la serie unidimensional (IR)

Para cada  $a \in \mathbb{G}$  llamemos  $N_a$  el nivel más alto para el que podemos obtener una descomposición de ondeletas para la serie unidimensional  $\mathcal{D}_a(t)$   $t \in \{1 \dots T\}$  (ec. 4.4) que corresponde a la neurona  $a$ . Llamemos  $N_0 = \max_{a \in \mathbb{G}} \{N_a\}$ , entonces para cada  $k \in \{1 \dots N_0\}$  podemos definir

$$M_k = \frac{1}{|\mathbb{G}_k| 2^{n-k}} \sum_{a \in \mathbb{G}_k, i \in \{1, 2^{n-k}\}} d_a(k, i) \quad (4.7)$$

Donde  $\mathbb{G}_k \subset \mathbb{G}$  es el conjunto de neuronas que aceptan una descomposición de nivel  $k$ ,  $d_a(k, i)$  es el  $i$ -ésimo coeficiente de nivel  $k$  en la descomposición de ondeletas de la serie  $\mathcal{D}_a(t)$  y  $n = \lceil \log_2(T) \rceil$ . Definimos entonces esta medida como

$$IR = \frac{1}{\max_{k \in \{1 \dots N_0\}} \{M_k\} - \min_{k \in \{1 \dots N_0\}} \{M_k\}} \quad (4.8)$$

Esta medida nos dice, en promedio sobre todas las neuronas, qué tan uniformemente están distribuidos los coeficientes a lo largo de los niveles. Si esta medida es baja, es decir, si los coeficientes promedio de cada nivel están contenidos en un intervalo grande, quiere decir que hay algún nivel cuyo coeficiente promedio es mucho

más grande que el coeficiente promedio de otro nivel. Si por el contrario, todos los niveles tienen coeficientes promedio parecidos, esta medida es alta.

Ésta es, entonces, una medida de la diversidad de escalas temporales. Puesto que el nivel de descomposición de ondeletas tiene que ver con que tanto nuestra serie de tiempo se parece a la ondeleta madre en cierta escala, podemos interpretarlo como qué tanto nuestra serie de tiempo parece registrar un fenómeno de duración igual a la duración de nuestra ondeleta madre escalada de cierta forma.

#### **4.4.3. Promedio por neurona de las varianzas (NMV)**

Una alta varianza de los coeficientes nos indica que no son muy parecidos. Esto por un lado nos sugiere que no son muchos los coeficientes cercanos a cero, y por otro lado nos sugiere que hay variedad en los valores absolutos de los coeficientes, es decir que hay variaciones de diversos tamaños en el espacio. Por otro lado una varianza baja nos indica que en el espacio casi todos los coeficientes son parecidos, es decir, la distancia entre el vector de entrada y el vector de pesos de la BMU tiene saltos aproximadamente del mismo tamaño. (Fig. 4-3). Esto es una medida de la diversidad de escalas espaciales, esta vez aplicada a las series unidimensionales. Llamamos a esta medida NMV y para calcularla obtenemos por cada neurona la varianza de todos los coeficientes de detalle de su DWT (sin importar el nivel) y luego sacamos el promedio, sobre todas las neuronas, de estas varianzas.

### **4.5. Resultados**

A continuación se presentan las gráficas (Figs. 4-4,4-5,4-6) obtenidas que comparan el desempeño del algoritmo (medido por medio del error topográfico) con las medidas de diversidad de escalas propuestas. Como puede verse, en las últimas dos (IR y NMV) hay una tendencia generalizada hacia abajo, es decir, en general, conforme la medida aumenta los errores topográficos obtenidos son menores.

Para ver esto un poco más claro tomemos, de todos los mapas hechos para cada conjunto de datos, los 400 (el 10%) que tengan el error topográfico más bajo y veamos



el histograma de cómo se distribuyen los mapas según su error topográfico (figs. 4-7,4-8 y 4-9). Como se observa en la medida IR, y en menor grado en la medida NMV, hay un corrimiento hacia la derecha del histograma, lo cual indica que en el mejor 10% de los mapas la medida tiende a ser alta.

Es, sin embargo, claro que una medida de error (TE) baja no está únicamente determinada por los fenómenos que hacen que ese mapeo tenga una medida (IR o NMV) alta. Aún así debe mencionarse que, como se observa en el cuadro 4.1 la correlación de estas dos medidas con el error topográfico es siempre negativa. Es importante notar que en el caso de la medida CL2, la correlación es siempre más baja que la correlación con las otras medidas, lo que sugiere que CL2 no captura tanta información respecto a procesos que sean relevantes para el desempeño del SOM. Es necesario un análisis adicional del comportamiento de la serie bidimensional para esclarecer su relación con el desempeño, pues hasta ahora solo se han observado las correlaciones de sus coeficientes de ondeletas de distintos niveles.

Por otro lado, vale la pena observar la correlación entre el error topográfico y los parámetros del SOM: los valores iniciales del parámetro de aprendizaje y del radio de la vecindad (Cuadro 4.2).

Medida	Conjunto de Datos	Correlación con TE
CL2	CODON	.0178
CL2	ION	.2096
CL2	IRIS	.1787
CL2	YEAST	-.0085
IR	CODON	-.4855
IR	ION	-.7896
IR	IRIS	-.6094
IR	YEAST	-.5865
NMV	CODON	-.4001
NMV	ION	-.2141
NMV	IRIS	-.5446
NMV	YEAST	-.4545

Cuadro 4.1: Correlaciones entre las medidas propuestas y el error topográfico.

Conjunto de Datos	Correlación con $\eta_{inic}$	Correlación con $U_{inic}$
CODON	.2710	-.2258
ION	.2127	-.3828
IRIS	.1760	-.2728
YEAST	.3901	-.3732

Cuadro 4.2: Correlación entre el TE y los parámetros del algoritmo  $\eta_{inic}$ , el valor inicial del parámetro de aprendizaje y  $U_{inic}$  el radio inicial de la vecindad

**Más que establecer que una medida alta implica un error topográfico bajo, los resultados aquí presentados indican que la probabilidad de obtener un buen mapa aumentan conforme aumentan las medidas IR y NMV de diversidad de escalas, en particular la medida IR .** Este enunciado constituye el resultado principal del trabajo.

Para explicar a qué nos referimos, observemos de nuevo el 10% de los mapas generados que tiene el menor error topográfico y observemos, para CL2, IR y NMV, cual es la media que se presentó. Comparemos esto con la media de CL2, IR y NMV para todos los mapas (Cuadro 4.3).

Conjunto de Datos	Medida	Media General	Media del mejor 10% de los mapas
CODON	CL2	1.6764	1.6149
CODON	IR	40.2677	59.8557
CODON	NMV	0.0732	0.0791
ION	CL2	1.2744	1.2475
ION	IR	1256.5	1712.1
ION	NMV	0.1998	.1767
IRIS	CL2	1.3031	1.2077
IRIS	IR	7.5173	9.9331
IRIS	NMV	.0130	.0177
YEAST	CL2	1.3906	1.3649
YEAST	IR	308.437	387.512
YEAST	NMV	0.0764	0.0841

Cuadro 4.3: Comparación de las medias para nuestras 3 medidas entre el 100% de los mapas generados y el 10% con menor TE

Como se ve en las dos últimas, en particular en la medida IR, la media es más alta para los 400 mejores mapas que para todos los 4,000 mapas. Esto indica que si tenemos un mapa bueno, es más probable que tenga una medida IR alta. Por otro lado, para

la medida CL2 ocurre lo contrario, lo cual sugiere que la probabilidad de encontrar un mapa con TE bajo disminuye conforme la correlación entre las coordenadas X y Y de la serie bidimensional aumenta. Esto último, a la luz de la baja correlación entre CL2 y TE, parece ser un fenómeno que se presenta en los mapas de error topográfico bajo, pues en el resto de los mapas no parece haber relación significativa entre CL2 y TE.

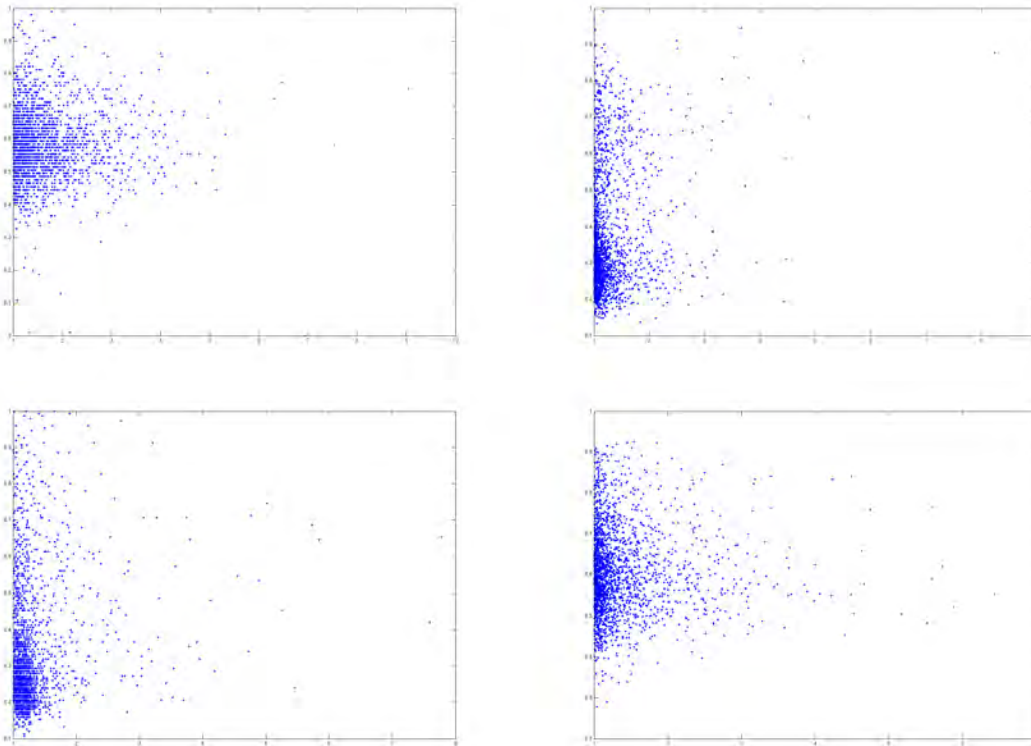


Figura 4-4: Inverso del rango de las correlaciones por nivel de los coeficientes de las coordenadas X y Y (2D) para las series (en el sentido de las manecillas desde arriba a la izquierda) Codón, Ion, Iris, Yeast, contra el error topográfico para 4,000 ejecuciones distintas. En el eje horizontal CL2, en el vertical TE.

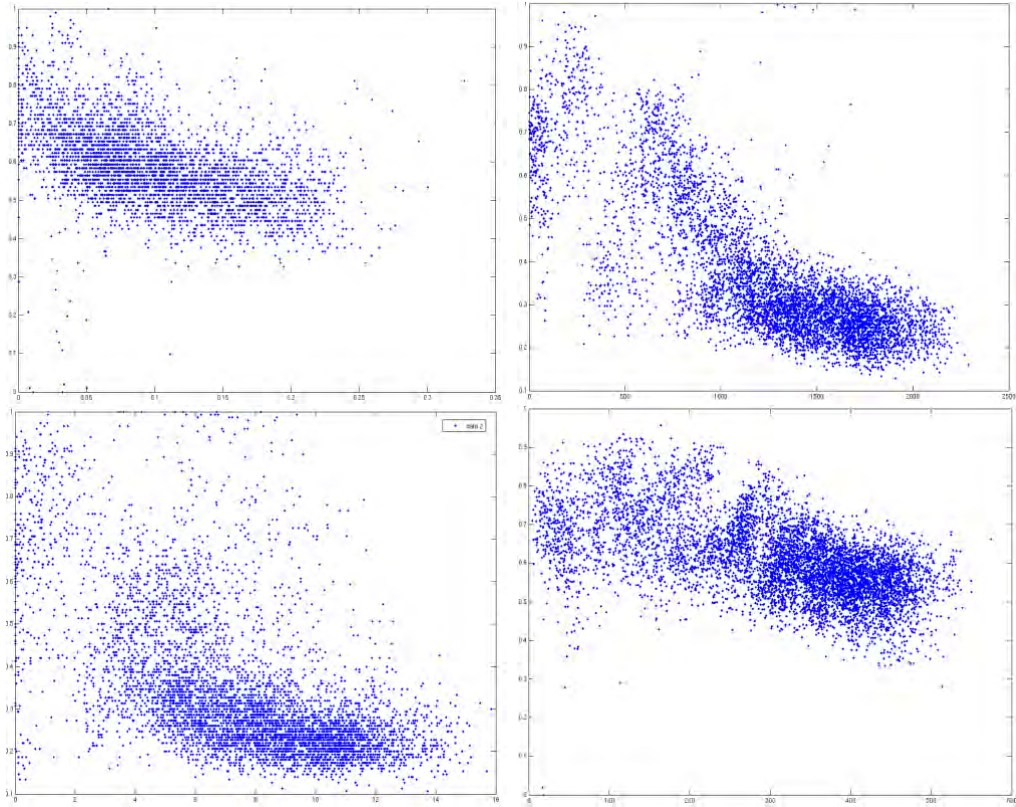


Figura 4-5: Inverso del rango de los promedios por nivel (1D) para las series (en el sentido de las manecillas desde arriba a la izquierda) Codón, Ion, Iris, Yeast, contra el error topográfico para 4,000 ejecuciones distintas. En el eje horizontal IR, en el vertical TE.

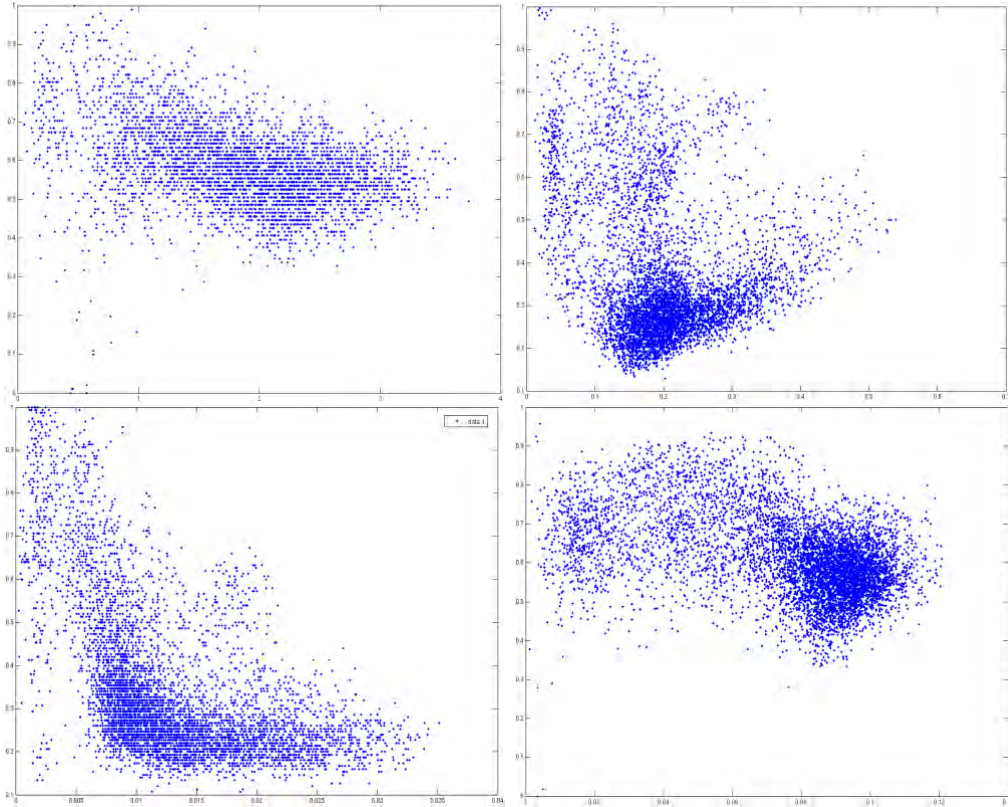


Figura 4-6: Promedio por neurona de las varianzas (1D) para las series (en el sentido de las manecillas desde arriba a la izquierda) Codón, Ion, Iris, Yeast, contra el error topográfico para 4,000 ejecuciones distintas. En el eje horizontal NMV, en el vertical TE.

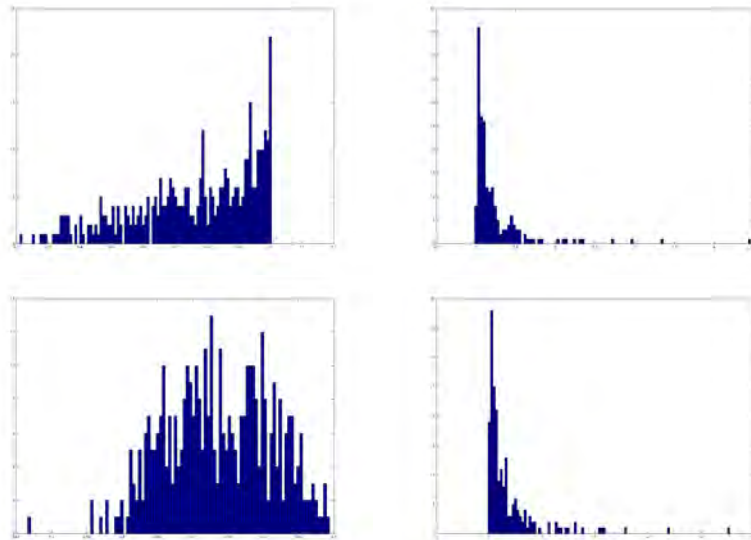


Figura 4-7: Histograma de frecuencias de aparición de la medida CL2 en los 400 mejores mapas para cada conjunto de datos. Gráficas correspondientes a las series (en el sentido de las manecillas desde arriba a la izquierda) Codón, Ion, Iris, Yeast. En el eje horizontal 100 subintervalos del intervalo que abarcan las medidas de error, en el eje vertical el número de mapas de entre los 400 mejores que cae en ese subintervalo.

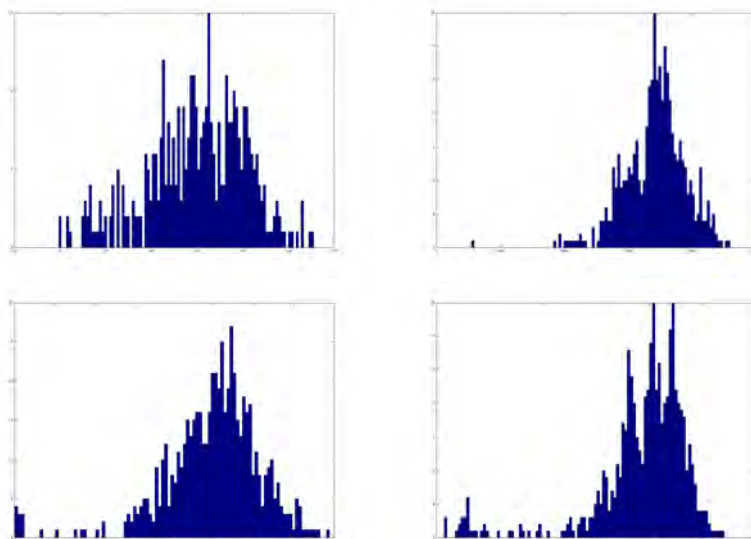


Figura 4-8: Histograma de frecuencias de aparición de la medida IR en los 400 mejores mapas para cada conjunto de datos. Gráficas correspondientes a las series (en el sentido de las manecillas desde arriba a la izquierda) Codón, Ion, Iris, Yeast. En el eje horizontal 100 subintervalos del intervalo que abarcan las medidas de error, en el eje vertical el número de mapas de entre los 400 mejores que cae en ese subintervalo.

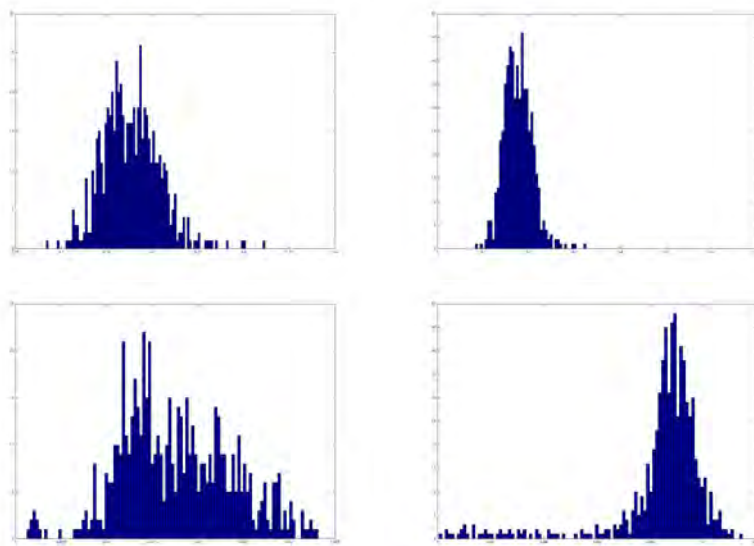


Figura 4-9: Histograma de frecuencias de aparición de la medida NMV en los 400 mejores mapas para cada conjunto de datos. Gráficas correspondientes a las series (en el sentido de las manecillas desde arriba a la izquierda) Codón, Ion, Iris, Yeast. En el eje horizontal 100 subintervalos del intervalo que abarcan las medidas de error, en el eje vertical el número de mapas de entre los 400 mejores que cae en ese subintervalo.



# Capítulo 5

## Discusión General

Los resultados obtenidos hasta ahora parecen sugerir que en efecto existe una relación entre la diversidad de las escalas de las variaciones durante la ejecución del SOM, en particular en la distancia entre el vector de pesos de la BMU y el estímulo que la activa, y su desempeño como herramienta en el análisis de cúmulos. Si bien esta relación no parece contundente, ya que las gráficas no son totalmente decrecientes, el hecho de que la distribución de los puntos (medida,error) se concentren alrededor de ciertas líneas o regiones es indicativo de que nuestras medidas IR y NMV de diversidad de escalas están, en efecto, capturando propiedades del comportamiento del SOM que son relevantes para su desempeño.

Por lo anterior creemos que intentar cuantificar el comportamiento del SOM, en particular en lo relacionado a la diversidad de escalas presentadas en las series unidimensionales aquí analizadas, puede rendir frutos importantes, sobre todo en lo que respecta a crear variantes del algoritmo que respondan a las tendencias observadas. Con esto queremos decir que este tipo de resultados puede servir como guía para proponer variantes del SOM que estén forzadas a tener el comportamiento que se observa para medidas altas y de esta forma aumentar la probabilidad de obtener buenos mapas. Una forma de hacer esto es monitorear constantemente las medidas (especialmente IR) que las series de tiempo unidimensionales presentan, y modificar el comportamiento del algoritmo adecuadamente, introduciendo un nuevo término en la ecuación 2.1 que rige la actualización de los pesos. Cabe mencionar que existe la

posibilidad, aún por verificarse, de que modificando el decrecimiento del umbral de la vecindad y el parámetro de aprendizaje se logren resultados similares.

Para que estas modificaciones se puedan hacer en tiempo real, es indispensable poder calcular las medidas conforme se ejecuta el algoritmo, y para ello la localidad temporal del análisis de ondeletas es indispensable.

Adicionalmente es posible que estas técnicas de cuantificación del comportamiento puedan aplicarse a otros sistemas que presentan auto organización para intentar comprenderlos mejor. Si bien se han reportado algunas características que presentan los sistemas que se auto organizan, por ejemplo[1], la caracterización completa de estos sistemas aún está lejos.

# Bibliografía

- [1] Bak P., Tang C., Wiesenfeld K. Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*. 59 (1987 ) 381-384
- [2] Bartumeus, F. et al. Helical Levy walks: Adjusting searching statistics to resource availability in microzooplankton *PNAS* 100 (2003) 12771-12775
- [3] Daubechies, I. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (1992) 331-371
- [4] DeRisi, J., Iyer, V., Brown, P. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 278(5338) (1997) 680-686
- [5] Erwin E., Obermayer K. & Schulten K. Self-Organizing Maps: Ordering, Convergence Properties and Energy Function. *Biological Cybernetics* 67. pp. 47-55 (1992)
- [6] Fichant, G., Gautier, C. Statistical method for predicting protein coding regions in nucleic acid sequences. *Bioinformatics*. Vol. 3 No. 4 (1987) 287-295
- [7] Fisher, R. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7 (1936) 179 - 188.
- [8] Folland, G. *Real Analysis Modern Techniques and Their Applications*. John Wiley & Sons (1999)
- [9] Folland, G. *Fourier analysis and its applications*. Brooks & Cole Publishing (1992)

- [10] Fort J. SOM's mathematics. *Neural Networks* no. 19 (2006) 812-816
- [11] Gardner, M. The fantastic combinations of John Conway's new solitaire game "life". *Scientific American* 223 (1970), 120-123.
- [12] Grantham R., Gautier C., Gouy M., Mercier R., A. Pave Codon catalog usage and the genome hypothesis. *Nucleic Acids Research*, Vol. 8, No. 1 (1980) 197
- [13] Grantham R. , Gautier C., Gouy M., Jacobzone M., Mercier R. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* no. 9(1) (1981) r43 - r74.
- [14] Haar A. Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.* 69 (1910)
- [15] Haken H. *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology* 3ra ed. Springer-Verlag (1983).
- [16] Haykin, S. *Neural Networks a Comprehensive Foundation*, 2nd ed. Prentice Hall 1999.
- [17] Ilachinski A. *Cellular Automata. A Discrete Universe.* World Scientific, (2001)
- [18] Kiviluoto K. Topology Preservation in Self-Organizing Maps. *Proceedings of International Conference on Neural Networks* (1996) 294-299
- [19] Koeing H., Greene, D. (2007) Prime Number Calculators en el sitio Game of Life News Bajado 14 de Marzo 2008 de [http://pentadecathlon.com/lifeNews/2007/12/prime\\_number\\_calculators.html](http://pentadecathlon.com/lifeNews/2007/12/prime_number_calculators.html)
- [20] Kohonen T. *Self-Organizing Maps.* Springer-Verlag (1995)
- [21] Kohonen T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43. pp. 59-69 (1982)
- [22] Kohonen T. Clustering, taxonomy, and topological maps of patterns. *Proceedings of the 6th Int. Conf. on Pattern Recognition.* pp. 114-128 (1982)

- [23] Korsnes R., et. al. Scaling in fracture and refreezing of sea ice *Physica A*. 331(2004) 291-296
- [24] Kraaijveld M, Mao J. Jain A. A non-linear projection method based on Kohonen's topology preserving maps. *Proceedings of the 11th Int. Conf. on Pattern Recognition*, pp. 41-45. (1992)
- [25] Kreiman G., Koch C. & Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience* 3. pp. 946-953 (2000)
- [26] Lepparanta, M. *The Drift of Sea Ice*. Springer (2005)
- [27] Liljeholm et al. Quantitative analysis of kernel properties in Kohonen's self-organizing map algorithm,: Gaussian and difference of Gaussians neighborhoods. *Neurocomputing* 44-46 pp. 515-520 (2002)
- [28] Lind D., Marcus B. *An introduction to symbolic dynamics and coding* Cambridge University Press (1995)
- [29] Mallat S. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Pattern Anal. and Machine Intell.*, 11, no. 7 (1989) 674-693
- [30] Moore D. *Estadística aplicada básica*. Bosch (1999)
- [31] Neme A., Miramontes P. A parameter in the SOM learning rule that incorporates activation frequency. *Proceedings of the Int. Conf. on Artificial Neural Networks 2006*. pp. 455-453. (2006)
- [32] Neme, A. et al. Decreasing neighborhood revisited in Self-organizing maps. Por publicarse.
- [33] Nesti, C. et al. Phylogeny inferred from codon usage pattern in 31 organisms. *Computer Applications in the Biosciences*, Vol. 11 No. 2 (1995) 167 - 171
- [34] Packard, N., Wolfram, S. *Journal of Statistical Physics*, 38 (1985) 901-946

- [35] Ramos-Fernandez, et al. Levy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). Behavioral Ecology and Sociobiology, 55 (2004) 223-230
- [36] Resnikoff H. Wavelet Analysis. The Scalable Structure of Information. Springer Verlag (1998)
- [37] Reynolds A.S Optimal scale-free searching strategies for the location of moving targets: New insights on visually cued mate location behaviour in insects Physics Letters A. 360 (2006) 224-227
- [38] Rosenblatt F. Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan Books (1961)
- [39] Shigeru N., Masaaki Y., Sadaki H.  $1/f$  fluctuation in the "Game of Life". Physica D. 118 (1998) 49-52
- [40] Sigillito, V., Wing, S., Hutton, L., & Baker, K. Classification of radar returns from the ionosphere using neural networks. Johns Hopkins APL Technical Digest, 10 (1989) 262-266.
- [41] Walnut D. An Introduction to Wavelet Analysis. Birkhauser (2001)
- [42] Ward, S. Genetic analysis of invasive plant populations at different spatial scales Biological Invasions. 8 (2006) 541-552
- [43] Weisstein E. (2000) Treasure Trove: The Life Cellular Automaton. Bajado 11 de Marzo 2008 de <http://www.ericweisstein.com/encyclopedias/life/>
- [44] Weisz N. et al. Tonotopic organization of the human auditory cortex probed with frequency-modulated tones. Hearing Research 191. pp. 49-58 (2004)
- [45] Wilshaw D. & von der Malsburg C. How patterned neural connections can be set up by self-organization. Proceedings of the Royal Society of London B 194. pp. 431-445 (1976)

- [46] Yamane S., Kajii S. & Kawano K. What facial features activate face neurons in the inferotemporal cortex of the monkey? *Journal Experimental Brain Research* 73. pp 209-214 (1988)