



**UNIVERSIDAD DE BUENOS AIRES**  
**Facultad de Ciencias Exactas y Naturales**  
**Departamento de Matemática**

**Tesis de Licenciatura**

**ANÁLISIS TOPOLÓGICO DE DATOS EN ATRACTORES CLIMÁTICOS**

**Luciana Salvagni**

**Director:** Pablo Amster

**Co-directoras:** Gisela Daniela Charó y Denisse Sciamarella

13 de Junio de 2023



# Índice general

<b>Índice general</b>	<b>1</b>
<b>Introducción</b>	<b>3</b>
<b>1. Conceptos básicos</b>	<b>5</b>
1.1. Sistemas no lineales . . . . .	5
1.2. Atractores climáticos . . . . .	7
1.3. Variedades enramadas . . . . .	8
1.4. Complejos celulares . . . . .	11
1.5. Series temporales . . . . .	13
1.6. Espectros singulares . . . . .	15
<b>2. Homologías persistentes</b>	<b>21</b>
2.1. Topología algebraica . . . . .	22
2.1.1. Homotopías . . . . .	22
2.1.2. Homologías . . . . .	23
2.2. Nervio de un cubrimiento . . . . .	27
2.3. Complejos Čech y Vietoris-Rips . . . . .	28
2.4. Homologías persistentes . . . . .	29
2.4.1. Filtraciones . . . . .	31
2.4.2. Diagramas de Persistencia . . . . .	32
<b>3. Análisis topológico de atractores climáticos</b>	<b>37</b>
3.1. Simulaciones con modelos climáticos . . . . .	39
3.2. Series temporales atmosféricas y oceánicas . . . . .	40
3.2.1. Pre-procesamiento . . . . .	41
3.2.2. Nubes de puntos . . . . .	44
3.3. Homologías persistentes y atractores climáticos . . . . .	46
3.3.1. Diagramas de persistencia . . . . .	49
3.3.2. Marcadores topológicos . . . . .	51
3.4. Discusión . . . . .	53
<b>4. Conclusiones y perspectivas</b>	<b>57</b>
<b>Bibliografía</b>	<b>61</b>



# Introducción

Las bases del análisis topológico para el estudio de los sistemas dinámicos se remontan a las contribuciones de Henri Poincaré, matemático, físico y filósofo de la ciencia, cuyos aportes tienen una enorme vigencia en distintos ámbitos de la matemática aplicada. Fue Poincaré quien advirtió, antes que nadie, el modo fundamental en que las soluciones de un sistema dinámico dependen de su topología.

La importancia de estos conceptos para las ciencias del clima, cuya definición no puede darse sin una comprensión adecuada del comportamiento a largo plazo de un sistema, no deja de aumentar frente a la emergencia que representa el cambio climático y el calentamiento global. El primer paso para explorar las propiedades de sistemas dinámicos como el clima terrestre consiste en identificar las distintas regiones del espacio de fases en las que evolucionan asintóticamente las trayectorias denominadas “atractores”.

Se ha demostrado que en un sistema climático existe la multiestabilidad. Cuando un sistema dinámico es multiestable, las soluciones para un valor dado de los parámetros internos pueden ser atraídas a diferentes regiones en el espacio de fases, para un mismo forzante y un nivel fijo de dióxido de carbono atmosférico. Esto significa que pueden co-existir varios atractores de características diferentes, y también significa que frente a ciertas perturbaciones, el sistema podría ‘migrar’ de un atractor a otro, cambiando radical e irreversiblemente. Estas crisis endógenas suelen denominarse ‘tipping points’ en dinámica del clima. Un ejemplo de tal cambio en el clima actual es la interrupción de la circulación de retorno en el Océano Atlántico.

Para estudiar el problema de la multiestabilidad en el clima, se utilizan configuraciones simplificadas en simulaciones numéricas basadas con modelos acoplados de atmósfera océano y hielo marino. Una de ellas considera un planeta de las dimensiones de la Tierra pero totalmente cubierto por agua. La búsqueda sistemática de atractores en un acuplaneta ha permitido encontrar un número variable de atractores coexistentes, cada uno de ellos asociados a un clima diferente, que va desde condiciones de bola de nieve hasta un estado caliente en el que el hielo marino desaparece por completo.

Estos atractores han sido caracterizados en términos de la circulación oceánica, el transporte de calor, la nubosidad y la distribución de la temperatura del aire en superficie, y estimando una lista de cantidades globales promediadas y componentes del balance energético para revelar las retroalimentaciones no lineales dominantes en cada atractor. Estas descripciones se apoyan en el análisis de componentes principales y en magnitudes métricas tales como la dimensión instantánea del atractor, sin examinar, no obstante, su topología.

La topología permite describir, identificar y clasificar un atractor según sus rasgos estructurales predominantes en el espacio de fases. A diferencia de otros invariantes de un sistema dinámico, como los exponentes de Lyapunov o las dimensiones fractales, la topolo-

gía permite dar cuenta del modo en que los procesos no lineales se combinan y se articulan para constituir un tipo determinado de atractor. Diferenciar topológicamente los atractores climáticos que coexisten en el espacio de fases, resulta pues de gran interés. Para hacerlo, es preciso realizar un análisis topológico del atractor climático, expresado como una nube de puntos en un espacio multidimensional.

El uso de herramientas de la topología algebraica ha tenido un éxito sorprendente en el análisis y reconocimiento de patrones a partir de nubes de puntos. El método de mayor popularidad se denomina homología persistente y tiene numerosas aplicaciones a problemas científicos tales como la detección de patologías a través del procesamiento de imágenes en medicina. En ciencias del clima, se ha comenzado a investigar si las homología persistente pueden ser utilizadas para detectar ríos atmosféricos, es decir, regiones de la atmósfera con intenso transporte de agua integrado verticalmente.

En el campo del análisis de series temporales, la información sobre patrones cuasi-periódicos y cambios en la dinámica de un sistema ha sido estudiada analizando las homología persistente de embebidos utilizados para reconstruir el espacio de fases a partir de series temporales multivariable. Los resultados han demostrado ser útiles, en algunos casos, para proponer índices topológicos que permiten detectar cambios dinámicos en el curso de una medición.

La descripción topológica que se obtiene con las homología persistente examina una multiplicidad de estructuras llamadas complejos celulares, que se construyen a partir de una única nube de puntos haciendo variar un parámetro relacionado con la distancia entre los puntos de la nube, denominado parámetro de filtrado. Ninguno de estos complejos es representativo por sí solo de la estructura topológica de un atractor en el espacio de fases. No obstante, el conjunto de propiedades topológicas de los complejos que persisten a la variación del parámetro de filtrado, provee marcadores topológicos que describen los rasgos primarios de la nube de puntos, y en consecuencia, también de un atractor bajo esa forma.

En esta tesis, realizamos la primera exploración de la posible utilidad de la técnica de homología persistente para estudiar la estructura de un atractor climático a partir de un conjunto de series temporales. El estudio concierne distintos atractores de un acuaplaneta acoplado, denominados respectivamente, estado caliente, estado templado, estado frío y cinturón de agua, a partir de resultados de simulaciones numéricas reportados en 2019. Se trabaja con series temporales asociadas a las siguientes variables escalares: temperatura del aire en superficie, humedad específica en la superficie, temperatura de la superficie del mar y precipitación.

La tesis se estructura en cuatro capítulos. El primero introduce definiciones básicas de la teoría de sistemas dinámicos y nociones generales de topología del caos, desde la perspectiva ofrecida por los grupos de homología. El segundo capítulo presenta la especificidad de las homología persistente en el análisis a partir de nubes de puntos. Un tercer capítulo presenta los datos de los cuatro atractores climáticos del acuaplaneta que se pretende caracterizar, junto con los diagramas de persistencia y marcadores topológicos. El último capítulo presenta las conclusiones y perspectivas que surgen de los resultados obtenidos.

# Capítulo 1

## Conceptos básicos

La primera frase de la novela Anna Karenina de León Tolstoi es: “Las familias felices son todas iguales; cada familia infeliz es infeliz a su manera”. Siguiendo al escritor ruso, Robert Gilmore y Marc Lefranc, fundadores del campo de investigación que hoy se conoce como ‘Topología del Caos’, nos dicen:

Los sistemas lineales son todos iguales; cada sistema no lineal es no lineal a su manera.

Y dicen a continuación:

Fue un descubrimiento muy feliz e impactante que hubiera estructuras en los sistemas no lineales que son siempre las mismas si se las mira de la manera correcta.

Ambas citas provienen del libro titulado ‘The Topology of Chaos: Alice in Stretch and Squeezeland’. (Gilmore and Lefranc, 2002)

En este capítulo, presentamos sucintamente los elementos fundamentales de teoría de sistemas dinámicos y de topología algebraica que convergen en este trabajo.

### 1.1. Sistemas no lineales

La teoría de sistemas dinámicos es un área de la matemática aplicada que procura describir la evolución temporal de un sistema. Podríamos incluso decir que se trata de una ‘matemática del tiempo’ (Smale, 1980). Cuando se emplean ecuaciones diferenciales ordinarias y el tiempo es una variable continua, la teoría se denomina sistemas dinámicos continuos.

Desde un punto de vista físico, los sistemas dinámicos continuos son una generalización de la mecánica clásica, y especialmente, del estudio de los movimientos de los cuerpos del sistema solar. A pesar de la gran elegancia y simplicidad de tales ecuaciones, la solución de problemas específicos resultó notablemente difícil hasta la llegada de Henri Poincaré (1854-1912), que unió análisis y geometría para desarrollar un enfoque cualitativo del estudio de las ecuaciones diferenciales.

En los sistemas de ecuaciones de primer orden que podemos escribir bajo la forma:

$$\frac{dx}{dt} = \dot{x} = f(x, t), \quad (1.1)$$

donde el tiempo  $t$  es la única variable independiente,  $x(t) \in \mathbb{R}^n$ , y  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^m$  es el campo vectorial o campo vector, decimos que el sistema es lineal si el campo vector es una

función lineal de las variables dependientes. En caso contrario, decimos que el sistema dinámico es no lineal.

A diferencia de un sistema no lineal, un sistema lineal puede desglosarse en problemas sencillos, cuya superposición permite, a posteriori, recuperar la solución del sistema original. En el caso de un sistema no lineal, en cambio, las partes cooperan, interfieren y se potencian, enriqueciendo la dinámica. No obstante, existen estructuras que permiten comprender en qué medida una solución de un sistema no lineal se parece a otra, si uno las representa ‘de la manera correcta’. Esa manera correcta de observar un sistema dinámico nos lleva a definir el espacio de fases.

### **Definición 1. Espacio de fases**

*El espacio constituido por las variables en términos de las cuales está dada la prescripción determinista que rige la dinámica del problema, se conoce como espacio de fases. En ese espacio, denominamos trayectoria a la curva que corresponde a seguir la evolución de una condición inicial (Mindlin, 2017).*

El concepto de espacio de fases fue desarrollado en el siglo XIX y de manera independiente por tres autores: Henri Poincaré, Ludwig Boltzmann y Willard Gibbs. Se trata de una noción fundamental en teoría de sistemas dinámicos, cuya clave está en asignar dimensiones coordinadas exclusivamente a las variables dependientes del tiempo, sin representar al tiempo como un eje coordinado adicional. Un sistema dinámico cuyo campo vector no depende explícitamente del tiempo se dice autónomo, y en ese caso, hay tantas ecuaciones de evolución como variables tiene el sistema y por lo tanto, la dimensión del espacio de fases coincide con el número de ecuaciones diferenciales ordinarias de primer orden. En este caso, cada punto representa unívocamente un estado del sistema.

### **Definición 2. Flujo**

*Se dice que el campo vectorial genera un flujo, cuando existe un difeomorfismo local  $\phi(t, x)$  para cada  $t$  que satisface la ecuación (1.1).*

Un punto fijo para un flujo en un sistema dinámico autónomo es un punto  $x_0 \in R^n$  tal que  $f(x_0) = 0$ . Estos puntos, también llamados puntos de equilibrio o ceros, son una clase importante de soluciones de una ecuación diferencial. Se dice que un punto fijo  $x_0$  es estable si la solución  $x(t)$  en su entorno permanece próxima a  $x_0$  durante un tiempo determinado. Un punto fijo se denomina inestable si no es estable. En algunos casos, la solución de un sistema dinámico puede ser periódica, lo que significa que las variables del sistema oscilan alrededor de un valor medio de manera repetitiva. La periodicidad se evidencia en el espacio de fases cuando se visitan repetidamente los mismos estados del sistema, es decir, los mismos puntos del espacio de fases. Si la serie es lo suficientemente larga (abarca un periodo de tiempo grande) y el intervalo de tiempo entre cada valor de la serie es lo suficientemente pequeño, el aspecto de la solución periódica será el de una curva cerrada. En algunos casos, si se deja evolucionar el sistema a partir de un estado inicial dado, es posible observar que la trayectoria solución, aun siendo aperiódica, queda finalmente ‘atrapada’ en una estructura cuya descripción será central para poner un poco de orden, en el desorden de la no linealidad.

Cuando se dispone de un sistema no autónomo, como sucede a menudo en los problemas de dinámica del clima, cabe examinar la posibilidad de re-escribir el sistema de ecuaciones



(1.1) de manera tal que la dependencia temporal sea absorbida por una o más variables adicionales. Con el número de variables dependientes del tiempo que se agregan al sistema para autonomizarlo, aumenta la dimensión del espacio de fases.

En el llamado problema de los tres cuerpos, Henri Poincaré ya había visto la analogía entre las soluciones que encontró y la “dependencia sensible sobre las condiciones iniciales” cuyo ejemplo paradigmático fue el del estudio del tiempo meteorológico. De hecho, en el Libro I, Cap. IV de Poincaré (1908), llamado ‘El Azar’, el polímata se pregunta:

“¿Por qué los meteorólogos tienen tanta dificultad para predecir el tiempo con certeza?  
¿Por qué las lluvias, las tempestades mismas parecen sobrevenir por casualidad, de modo que a muchas personas les parece muy natural rezar para que llueva o haga sol, cuando les parecería ridículo rezar por un eclipse?”

Y nota especialmente que “puede ocurrir que pequeñas diferencias en las condiciones iniciales produzcan otras muy grandes en los fenómenos finales.”

Precursor de lo que hoy constituye la teoría del caos, Poincaré también fue pionero en diversos aspectos que hoy cubre la rama de la matemática dedicada al estudio de aquellas propiedades de los cuerpos geométricos que permanecen inalteradas por transformaciones continuas, advirtiendo de qué modo las propiedades de un sistema dinámico dependen de lo que hoy denominamos topología. Citémoslo una vez más:

“Existe una ciencia llamada *Analysis Situs*, cuyo objeto es el estudio de las relaciones de posición de los diversos elementos de una figura haciendo abstracción de sus tamaños. Esta geometría es puramente cualitativa; sus teoremas serían ciertos si las figuras, en lugar de ser exactas, fueran burdamente imitadas por un niño. También podemos hacer un *Analysis Situs* con más de tres dimensiones. La importancia del *Analysis Situs* es enorme y no me canso de resaltarlo.”

## 1.2. Atractores climáticos

La meteorología provee el ejemplo paradigmático de la capacidad de un sistema (la atmósfera) para cambiar de forma radical e imprevisible. Los científicos alentaron durante mucho tiempo la esperanza de que el clima, definido como una media a largo plazo, fuera más estable: bastaba que las fluctuaciones diarias se anularan unas a otras con el correr del tiempo, para que el sistema atmosférico, se auto-estabilizara.

Hoy sabemos que el clima de nuestro planeta es un sistema fuera de equilibrio que, bajo el efecto de una distribución no homogénea de la radiación solar en su frontera, evoluciona hacia estados estadísticamente estacionarios. Las interacciones no lineales entre los principales componentes del sistema climático, es decir la atmósfera, el océano, la criosfera y la biosfera aumentan la complejidad de su estudio. Por esta razón, se ha desarrollado históricamente una jerarquía de modelos cada vez más exhaustivos, que van desde los modelos de balance energético hasta los modelos de circulación general más realistas.

En esta tesis, estudiaremos una configuración idealizada denominada acuaplaneta, en la que el planeta está homogéneamente cubierto por un océano de tres mil metros de profundidad conservando el acoplamiento completo entre la atmósfera, el océano y el hielo marino. De este modo, aunque la complejidad del sistema se reduce con respecto al sistema climático real, se siguen teniendo en cuenta las principales interacciones no lineales entre sus componentes.

Bajo la acción de un forzante externo constante y a valores fijos de los parámetros internos, las soluciones de un sistema dinámico son atraídas hacia regiones estables del espacio de fases, denominadas atractores. Un atractor es el conjunto cerrado invariante mínimo que atrae a un conjunto abierto de condiciones iniciales que convergen hacia él a medida que el tiempo tiende a infinito. El mayor de estos conjuntos se denomina cuenca de atracción. Encontrar los atractores es el primer paso en el análisis de la dinámica no perturbada. En los límites de la cuenca, la dinámica es altamente no lineal, y pequeñas perturbaciones dan lugar a cambios bruscos y potencialmente irreversibles que corresponden al paso de un atractor a otro.

La caracterización de los atractores climáticos ayuda a predecir la respuesta del sistema ante perturbaciones. Actualmente se está realizando un gran esfuerzo para comprender las propiedades de los atractores climáticos, con un especial interés por las transiciones inducidas por el ruido. Nuestra comprensión de la interacción del ruido con la no linealidad de un sistema dinámico es aún relativamente limitada, y es fuente de estudios que conciernen incluso los sistemas no lineales de baja dimensión, como el atractor aleatorio denominado LoRA (Lorenz Random Attractor) (Charó et al., 2021).

Distintos estudios han logrado identificar un número reducido de atractores en un acua-planeta que coexisten bajo el mismo forzante externo, representado en las simulaciones numéricas por valores fijos de irradiación solar y contenido atmosférico de dióxido de carbono. La coexistencia de una pluralidad de estados estables para un mismo forzante externo pone en evidencia la complejidad del modelo más sencillo de un sistema climático acoplado que es posible concebir.

Cada uno de los atractores ha sido descrito estimando una lista de cantidades globales promediadas para intentar diferenciarlos entre sí. Los estados frío/caliente pueden corresponder a los ciclos glaciares/interglaciares de los últimos tres millones de años (Ferreira et al., 2018). Del mismo modo, además del ciclo glacial/interglacial, se propone otro ciclo para el clima actual que incluye un estado caliente hacia el que se aproxima el Sistema Tierra bajo el efecto del calentamiento global<sup>1</sup>. Resulta en consecuencia de sumo interés cuantificar la existencia de ese posible cambio de un estado templado a uno caliente partiendo de una configuración acuaplanetaria simple.

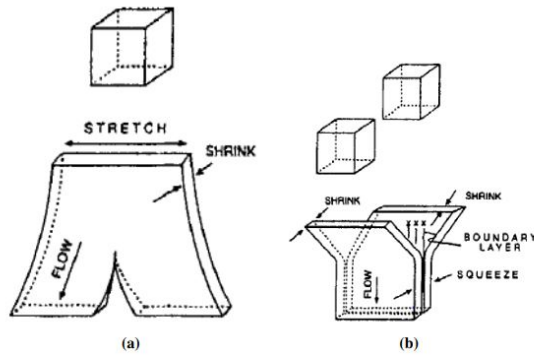
Cada atractor es el resultado de la competencia entre diferentes mecanismos no lineales. Pero, ¿qué herramientas ofrece la teoría de sistemas dinámicos no lineales para describir tales mecanismos?

### 1.3. Variedades enramadas

El estudio de la estructura topológica de un flujo en el espacio de fases cobra impulso a partir de los trabajos de Robert Gilmore. Este especialista en teoría de grupos y dinámica no lineal explica que los invariantes que se utilizan para caracterizar un conjunto de trayectorias de un sistema en régimen caótico (como la dimensión fractal o los exponentes de Lyapunov) no nos proporcionan una forma de distinguir entre los distintos tipos de atractores: se necesita una herramienta de naturaleza diferente para crear un diccionario de procesos y mecanismos subyacentes a un sistema no lineal. Esta herramienta resulta ser de naturaleza topológica y el objeto de dicha descripción lo que hoy denominamos ‘variedad

---

<sup>1</sup>Este resultado se basa en un análisis cualitativo Steffen et al. (2018), que comienza a ser cuantificado por primera vez en Brunetti et al. (2019) partiendo de una configuración acuaplanetaria simple.



**Figura 1.1:** (a) Un volumen en el espacio de fases que por efecto del estiramiento y compresión da lugar a la formación de dos ramas, (b) dos volúmenes que se acercan por efecto del plegado del flujo, formándose una sola rama. En la Figura 1.1 se puede observar los efectos de estos mecanismos. Figura extraída de Gilmore and Lefranc (2002).

enramada’.

**Definición 3. Variedad enramada**

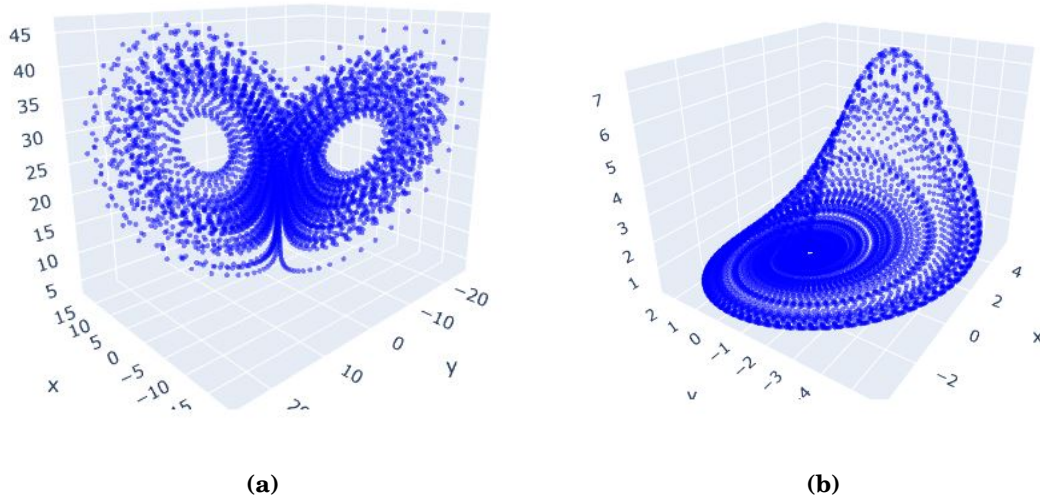
Una variedad  $n$ -dimensional es un espacio topológico tal que cada punto tiene una vecindad topológicamente equivalente a un disco abierto  $n$ -dimensional. Si para dos puntos cualesquiera existen entornos disjuntos, se dice que la variedad es Hausdorff. Denominamos variedad enramada a aquella variedad a la que no se le exige que cumpla la propiedad de Hausdorff.

El entorno de la solución a largo plazo (suprimido el transitorio) de un sistema dinámico determinístico cobra la forma de una variedad enramada que no cambia, que constituye un invariante del sistema dinámico. Esta variedad enramada es una estructura cuyas ramas marcan los sitios en donde las trayectorias se separan para partir hacia regiones diferenciadas del espacio de fases – Figura 1.1. La propiedad de Hausdorff no se cumple precisamente en los lugares que describen el estiramiento y compresión del flujo que permite que haya destinos diferentes para las trayectorias. La cantidad y ubicación relativa de las ramas de una variedad enramada dan lugar a distintos tipos de comportamiento caótico, permitiendo distinguir entre distintos tipos de atractores.

“Las variedades enramadas son construcciones útiles para distinguir entre los distintos mecanismos que generan atractores. La equivalencia topológica entre las variedades enramadas se produce por isotopía. Dos objetos son isotópicos si es posible moldear uno en el otro sin cortar ni pegar. Como resultado, identificar la variedad enramada que describe un atractor es una poderosa herramienta para distinguir una (clase de) atractores de otra.”

Por ejemplo, el atractor de Rössler, regido por este conjunto de tres ecuaciones diferenciales ordinarias para  $a = 0,43295, b = 2, c = 4$ :

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= ay + x \\ \dot{z} &= b + z(x - c), \end{aligned}$$



**Figura 1.2:** Los atractores de Lorenz y de Rössler tienen la misma dimensión fractal, pero están asociadas a variedades enramadas topológicamente diferentes.

y el atractor de Lorenz con  $\rho = 28, \sigma = 10, \beta = \frac{8}{3}$ :

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z.\end{aligned}$$

tienen el mismo valor de dimensión fractal ( $d = 2,61$ ) a pesar de tener variedades enramadas muy distintas. Es sencillo ver que el atractor de Rössler tiene un punto fijo inestable del cual las trayectorias se alejan en espiral (foco repulsor), mientras que el atractor de Lorenz tiene dos focos repulsivos, situados en el centro de cada una de las alas de la mariposa. El modo en que el flujo pasa de un ala a la otra en el atractor de Lorenz o la manera en la que la espiral del atractor de Rössler se reinserta después de una torsión en la coordenada  $z$  caracterizan y distinguen las variedades enramadas de estos atractores. Las nubes de puntos correspondientes a los dos atractores caóticos se muestran en la Figura 1.2.

Las variedades enramadas de estos dos atractores son invariantes del sistema determinístico que genera las nubes de puntos de la 1.2. En la Figura 1.3 se esbozan distintas deformaciones geométricas del atractor de Lorenz utilizando transformaciones de coordenadas. La geometría de la nube de puntos cambia, pero el número de ramas y el modo en que se reinyectan después de separarse, en la misma variedad, permanecen intactas (Ghil and Sciamarella, 2023). La topología de la variedad enramada que ‘soporta’ un flujo determinístico puede en consecuencia caracterizarse y utilizarse para reconocer un determinado tipo de atractor, independientemente de la forma geométrica particular de la nube de puntos que se utiliza para visualizarla.

¿Existe la posibilidad de que la topología de una variedad enramada cambie en el tiempo? La topología es un invariante si el sistema es determinístico. Un sistema dinámico es determinístico si para un estado actual hay un único futuro. Es estocástico si hay una distribución de probabilidades de los posibles estados futuros. En esta tesis, no nos ocuparemos de los sistemas dinámicos estocásticos, pero tendremos en cuenta que la topología de una variedad enramada puede alterarse si el sistema dinámico no es determinístico. También es posible observar cambios en la variedad enramada de un sistema determinístico, antes y después de una bifurcación. Una bifurcación se produce cuando un pequeño cambio



**Figura 1.3:** Nubes de puntos asociadas a diferentes representaciones geométricas del atractor de Lorenz. Figura extraída de Ghil and Sciamarella (2023).

suave realizado en los valores de los parámetros (a menudo denominados parámetros de control) provoca un cambio “cualitativo” repentino en su comportamiento. En los términos de esta tesis, una bifurcación se traduce en un cambio topológico en la variedad enramada, como puede ser la desaparición de una rama o la aparición de una torsión antes ausente.

Dada una nube de puntos, como la de las Figura 1.2 o 1.3, ¿es posible extraer la topología de la variedad enramada que las soporta? El problema ante el cual estamos presentes es el del cómputo de las propiedades topológicas de la variedad enramada que subyace una nube de puntos en un espacio de fases.

## 1.4. Complejos celulares

La topología algebraica propone instrumentos para transformar un problema de topología en un problema de álgebra. Le debemos a un matemático británico nacido en la India, J. H. C. Whitehead, el concepto de complejo celular, que utilizaremos en esta tesis para convertir una variedad enramada en un objeto sobre el cual puedan hacerse operaciones algebraicas.

Dado un atractor bajo la forma de una nube de puntos en el espacio de fases de dimensión  $n$ , es posible aproximar y así obtener las propiedades topológicas de la variedad enramada subyacente, agrupando los puntos en celdas.

### Definición 4. Celda

*Una  $m$ -celda es un conjunto cuyo interior es homeomorfo al disco  $m$ -dimensional  $D_m = \{x \in R^m : \|x\| < 1\}$  con la propiedad adicional de que su límite o frontera debe dividirse en un número finito de celdas de menor dimensión, llamadas caras de la  $m$ -celda. Escribimos  $\sigma < \tau$  si  $\sigma$  es una cara de  $\tau$ .*

(i) Una celda de dimensión 0 es un punto  $A$ .

(ii) Una 1-celda (o celda de dimensión 1) es un segmento de recta  $AB$ , y  $A < AB$ ,  $B < AB$ .

(iii) Una 2-celda (o celda de dimensión 2) es un polígono. Si  $\sigma = \triangle ABC$ , entonces  $AB$ ,  $BC$  y  $AC < \sigma$ . Obsérvese que como  $A < AB$  y  $AB < \sigma$  entonces  $A < \sigma$ .

(iv) Una 3-celda es un poliedro sólido (a menudo un tetraedro), con polígonos, aristas y vértices como caras.

Dada una nube de puntos en el espacio de fases, un método que aproxima la variedad enramada como un conjunto de celdas, en un sentido bien preciso, se encuentra detallado en (Sciamarella and Mindlin, 1999, 2001), y que resumiremos a continuación.

Sea  $n$  la dimensión del espacio de fases, y sea  $\mathcal{N}$  una nube de puntos representativa de un atractor que yace en una variedad enramada de dimensión local  $m$ . En el caso del atractor de Lorenz, por ejemplo, tenemos  $n = 3$  y  $m = 2$ . El hecho de que los puntos de  $n$  coordenadas yazgan en una variedad enramada puede utilizarse para agrupar a los puntos en  $m$ -celdas que vayan cubriendo progresivamente la variedad. El subconjunto de puntos asociado a cada una de estas  $m$ -celdas debe aproximar un  $m$ -disco. Sea  $S(x_0) = \{x_i, i = 1, \dots, n_c\}$  un subconjunto de  $n_c$  puntos de la nube  $\mathcal{N}$ , en torno a un punto  $x_0$ . Para verificar si  $S(x_0)$  puede ser aproximado por una  $m$ -celda, como  $\gamma_1$  en la Figura 1.4, se analiza si los puntos de  $S$  que pertenecen a una bola con centro  $x_0$  y radio  $r$  aproximan localmente a un disco  $m$ -dimensional. La distribución de puntos en  $S(x_0)$  puede analizarse construyendo la matriz  $X \in \mathbb{R}^{n_c \times n}$ :

$$X_{i,j} = \frac{1}{n_c^{1/2}}(x_{i,j} - x_{0,j}).$$

donde  $n_c$  es el número de puntos que se pretende agrupar en una sola  $m$ -celda. Los autovectores de  $X^T X$  se conocen con el nombre de vectores singulares de  $X$ , y a la raíz cuadrada de cada autovalor de  $X^T X$  se la denomina valor singular de  $X$ . Los vectores singulares proveen un sistema de coordenadas local centrado en  $x_0$ , y sus valores singulares describen la distribución de los puntos en torno a  $x_0$ . Si  $S(x_0)$  aproxima un disco de dimensión  $m$  en  $\mathbb{R}^n$ , deben existir  $m$  valores singulares de  $X$  que son función lineal del número de puntos en  $S(x_0)$  ordenados en una secuencia de distancia creciente a  $x_0$  (Muldoon et al., 1993). Los valores singulares restantes  $(n - m)$  medirán la desviación del espacio tangente al disco y escalarán como  $r^\ell$  con  $\ell \geq 2$ . Si el conjunto satisface este requerimiento, se le asigna a  $S(x_0)$  una  $m$ -celda que, siguiendo el criterio en (Sciamarella and Mindlin, 1999), compartirá caras con celdas vecinas para formar un complejo:

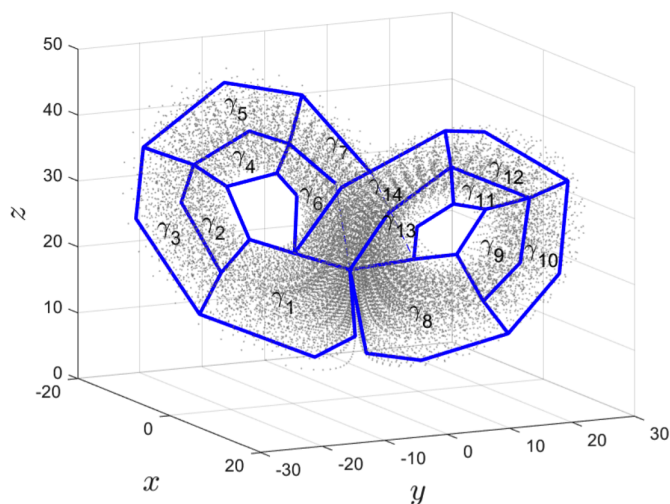
### **Definición 5. Complejo**

Un complejo  $K$  es un conjunto finito de celdas,  $K = \cup\{\sigma : \sigma \text{ es una celda}\}$  tal que:

- (i) si  $\sigma$  es una celda de  $K$ , entonces todas las caras de  $\sigma$  son elementos de  $K$ ;
- (ii) si  $\sigma$  y  $\tau$  son dos celdas distintas de  $K$ , entonces  $Int(\sigma) \cap Int(\tau) = \emptyset$ , donde  $Int(\sigma)$  denota el interior de  $\sigma$ .

La dimensión de  $K$  es la dimensión de su celda de mayor dimensión.

Un complejo de celdas construido de este modo sobre la nube de puntos de un atractor puede considerarse un topológicamente fiel a la variedad enramada que estructura el atractor. Dado que el método de construcción de un complejo de estas características utiliza la propiedad de que hay subconjuntos de puntos que aproximan localmente una variedad de dimensión  $m \leq n$ , llamamos ‘BraMAH’ (*Branched Manifold Analysis through Homologies*) a un complejo construido de este modo. La Figura 1.4 muestra un complejo BraMAH para la nube de puntos del atractor de Lorenz que, siguiendo este procedimiento, da lugar a catorce 2-celdas  $\gamma_i$  ( $i = 1, \dots, 14$ ).



**Figura 1.4:** Complejo celular del atractor de Lorenz construido con el método detallado en (Sciama-rella and Mindlin, 2001). Notemos que el espacio de fases es de dimensión  $n = 3$ , mientras que la dimensión local de la variedad enramada que coincide por construcción con la dimensión del complejo es de dimensión  $m = 2$ . Fuente: Charó et al. (2022).

Este no es, por cierto, el único modo de construir complejos celulares a partir de una nube de puntos, tal como se verá en el capítulo siguiente. Los métodos más ampliamente difundidos para construir complejos a partir de nubes de puntos, no utilizan la noción de variedad enramada y pueden aplicarse a cualquier nube de puntos, sin ningún requerimiento particular. La nube de puntos puede no estar asociada a una variedad enramada, tampoco a un flujo.

## 1.5. Series temporales

En ciencias del clima, pero también en mecánica de fluidos, y en muchas otras áreas del conocimiento científico, se dispone de series temporales, es decir, de un conjunto de medidas secuenciales de una o varias variables de un sistema dinámico a lo largo del tiempo, sin disponer del conjunto de ecuaciones diferenciales que podrían tener como solución a dicha serie temporal.

No obstante, existen un conjunto de técnicas agrupadas bajo el nombre de ‘embedología’ que permiten obtener, a partir de una o varias series temporales, una nube de puntos en un espacio de fases ‘reconstruido’, cuyas propiedades topológicas deseamos describir.

De las numerosas descripciones de este conjunto de métodos que nos permiten reconstruir el espacio de fases a partir de series temporales  $x(t)$ , seguiremos a (Natiello et al., 2007). Supongamos que en lugar de estudiar la serie

$$S_1 = \{x_i, i = 1, 2, \dots\}$$

construimos el siguiente conjunto de arreglos bidimensionales:

$$S_2 = \left\{ \left( x_i, \frac{x_i - x_{i-1}}{\Delta t} \right), i = 1, 2, \dots \right\}.$$

Si el espacio de fases fuese de dimensión 2, y la variable representara la evolución de un sistema cuyo comportamiento fuera estrictamente periódico, el gráfico en un espacio  $\mathbb{R}^2$  de

la serie temporal  $S_2$  dará lugar a un conjunto de puntos que describen una curva cerrada.

Para reconstruir el espacio de fases en dimensión  $n \geq 3$  a partir de un conjunto de series temporales  $x(t)$ , es necesario construir vectores  $s(t)$  con  $n$  componentes a partir del escalar  $x(t)$ . Una de las estrategias es utilizar embebidos por retardo temporal.

Sea  $n$  la dimensión del espacio de fases, consideremos el siguiente mapeo para construir embebidos  $s(t)$  usando el observable  $x(t)$ :

$$x(t) \rightarrow s(t) = (s_1(t), \dots, s_n(t)) \quad (1.2)$$

donde  $s_k(t) = x(t - \tau_k)$ ,  $k = 1, \dots, n$  y  $\tau_k \in \mathbb{R}_{>0}$ . Los parámetros  $\tau_k$  se denominan retardos temporales, usualmente se los considera equiespaciados ( $\tau_k = k \tau$ ). Para el caso de una serie temporal discreta  $x(i)$  se tiene:

$$x(i) \rightarrow s(i) = (s_1(i), \dots, s_n(i))$$

donde  $s_k(i) = x(i - (k - 1) \tau)$ ,  $k = 1, \dots, n$  y  $\tau \in \mathbb{N}$ . Este mapeo se denomina embebido por retardo temporal.

La topología de la variedad enramada que resulta de la reconstrucción dependerá del valor asignado al retardo temporal y del número de dimensiones utilizadas para reconstruir el espacio de fases (Gilmore, 1998). Los embebidos con retardo temporal requieren, en consecuencia, estrategias adicionales para determinar la mínima dimensión  $n_{min}$  del embebido; y el valor óptimo del retardo  $\tau$ .

Para determinar el óptimo valor de  $\tau$ , se apela a un método denominado ‘Información Mutua Media’ (AMI, *Average Mutual Information* en inglés) (Shannon, 1948). La información mutua media es un método que puede considerarse como una generalización no lineal de la función de autocorrelación entre dos variables. En el caso de los embebidos por retardos, AMI busca cuantificar la independencia entre una serie temporal  $x(t)$  y una versión desplazada de la misma  $x(t + \tau)$  con el objetivo de encontrar un  $\tau$  que verifique que las coordenadas  $(x(t), x(t + \tau))$  sean tan independientes entre sí como sea posible (Fraser and Swinney, 1986). Con el  $\tau$  hallado podremos encontrar el mejor embebido por retardos temporales que nos permite reconstruir el espacio de fases.

Para aplicar el método AMI a series temporales se debe crear un histograma que muestre la distribución de  $x(t)$  en intervalos equiespaciados dentro del rango de la variable. Sea  $p_i$  la probabilidad de que  $x(t)$  esté en la ubicación  $i$  del histograma, y  $p_{ij}(\tau)$  la probabilidad de que  $x(t)$  esté en la ubicación  $i$  y  $x(t + \tau)$  esté en la ubicación  $j$ . Definimos la información mutua  $I(x(t), x(t + \tau))$  como:

$$I(x(t), x(t + \tau)) = \sum_{i,j} p_{ij}(\tau) \log \left( \frac{p_{ij}(\tau)}{p_i p_j} \right). \quad (1.3)$$

Para que las coordenadas del embebido sean tan independientes como sea posible se establece como el valor óptimo de  $\tau$  la posición del primer mínimo de  $I(x(t), x(t + \tau))$ . Esto significa que la primera coordenada del embebido  $x(t)$  será máximamente independiente de la segunda coordenada  $x(t + \tau)$ , y así sucesivamente. El retardo temporal adecuado es el intervalo de tiempo para el cual existe un mínimo local en la información mutua media. En el caso que de la función AMI no tenga un mínimo local pero sea una función monótona decreciente de  $\tau$ , el valor de  $\tau$  se toma como el lapso de tiempo en el que se ha producido la primera disminución sustancial en la información mutua media.



Para determinar la mínima dimensión  $n_{min}$  del embebido se puede utilizar el método de los falsos vecinos más cercanos (FNN, False Nearest Neighbors) (Kennel et al., 1992). La idea principal es examinar cómo cambia el número de vecinos de un punto a lo largo de la trayectoria de una señal al aumentar la dimensión del embebido. En una dimensión demasiado baja, muchos de los vecinos serán falsos, pero en una dimensión adecuada o superior, los vecinos serán reales. Los falsos vecinos aparecerán como tales porque la estructura geométrica del atractor se ha proyectado hacia un espacio de menor dimensión, generando auto-intersecciones entre trayectorias que no deberían darse en un sistema determinístico donde para un estado del sistema, existe un único futuro. Aumentar la dimensión del embebido permite deshacer estas auto-intersecciones. Es en el pasaje de una dimensión  $d$  a la siguiente ( $d + 1$ ) que es posible distinguir entre los verdaderos y los falsos vecinos. Sea  $s^r(n)$  el vecino  $r$  de  $s(n)$  y  $R_d(n, r)$  la distancia entre ellos:

$$R_d^2(n, r) = \sum_{k=0}^{d-1} [x(n + k\tau) - x^r(n + k\tau)]^2.$$

El criterio de FNN designa como falso vecino más cercano a cualquier vecino para el que sea válido lo siguiente:

$$\left[ \frac{R_{d+1}^2(n, r) - R_d^2(n, r)}{R_d^2(n, r)} \right]^{\frac{1}{2}} > R_{tol},$$

donde  $R_{tol}$  es el umbral de la tolerancia en términos de distancia. Valores muy altos de  $R_{tol}$  resultarán en una subestimación del número de FNN, mientras que valores muy bajos de  $R_{tol}$  identificarán demasiados falsos vecinos, en particular cuando los puntos del atractor son escasos y la nube es poco densa. Elegimos un valor para  $n_{min}$  en el punto en el que el número de FNN cae a 0, o los embebidos posteriores no cambian el número de FNN o el punto justo antes del cual el número de FNN empieza a aumentar de nuevo.

## 1.6. Espectros singulares

En el caso de un fluido en movimiento, como la atmósfera o el océano de un acuaplaneta, es frecuente disponer de series temporales de variables medidas en un punto de grilla. Lo mismo ocurre en experimentos fluidodinámicos con sondas, como la velocimetría Laser Doppler, que permite medir la evolución de un componente de la velocidad en un punto de un fluido transparente o semi-transparente (Tuerke et al., 2017). ¿Qué problemas se plantean en la reconstrucción de un espacio de fases a partir de un conjunto de series temporales de variables distintas, como la temperatura y la humedad, tomadas por ejemplo en distintas estaciones meteorológicas?

Durante las tres últimas décadas, el análisis de espectros singulares (SSA, Singular Spectrum Analysis, en inglés) y su versión multicanal (M-SSA) han resultado muy populares en diversos campos de las geociencias. El nombre de la técnica se refiere al espectro de valores singulares que se obtienen de una matriz de covarianza, y no directamente a una descomposición en el dominio de la frecuencia. Esta técnica permite, dada una serie temporal, escribir esa serie como una suma de componentes, es decir, de funciones del tiempo que la aproximan. Las componentes principales, relacionadas con los valores singulares más altos, permiten reconstruir un ‘esqueleto’ robusto de la dinámica del sistema reduciendo la dimensión del espacio de fases correspondiente al conjunto inicial de datos y proporcionando así una visión simplificada de los mecanismos presentes (Groth and Ghil, 2011).

Algunos autores como (Feliks et al., 2010) han propuesto la aplicación de M-SSA a pares de variables climáticas con el objeto de determinar la influencia mutua y la sincronización entre estas variables. M-SSA permite reconstruir el comportamiento dinámico compartido por las variables, en términos de modos oscilatorios comunes. La técnica de M-SSA será utilizada en nuestro trabajo para lograr reconstrucciones simplificadas del espacio de fases a partir de conjuntos de series temporales, que permitan disminuir la dimensión del espacio de fases reconstruido, sin por ello dejar de captar los ingredientes principales de los mecanismos presentes.

Los autores de (Broomhead and King, 1986) introdujeron las descomposiciones espectrales de series temporales en el análisis de los sistemas dinámicos. Aquí resumimos la técnica de M-SSA, que será empleada para el análisis de las distintas variables medidas para estudiar los atractores climáticos.

Sea  $x_d(t)$  una serie temporal de longitud  $N$  y  $d \in \mathbb{N}$  con  $d \leq N$ , también llamada canal, el punto de partida de M-SSA es embeber cada canal en un espacio de fases de  $M$  dimensiones utilizando retardos temporales.

Sea  $\mathbf{X}_d$  la matriz de trayectoria del canal  $d \leq D$ :

$$\mathbf{X}_d = \begin{bmatrix} x_d(1) & x_d(2) & \dots & x_d(M) \\ x_d(2) & x_d(3) & \dots & x_d(M+1) \\ \vdots & \vdots & \vdots & \\ x_d(K) & x_d(K+1) & \dots & x_d(N) \end{bmatrix}$$

donde  $K = N - M + 1$ . Se define a  $\mathbf{X}$  como la matriz de trayectoria multicanal con  $N - M + 1$  filas y  $D \times M$  columnas:

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_D).$$

La elección de  $M$  está basada en un compromiso entre dos consideraciones: la cantidad de información extraída y el grado de confianza estadística de la misma. La primera requiere un  $M$  tan grande como sea posible, pues cuanto más grande es  $M$ , más información se puede obtener, y la segunda consideración requiere tener tantas repeticiones como sea posible del patrón de interés, es decir un valor alto de  $N/M$ .

El algoritmo M-SSA realiza una descomposición en valores singulares (SVD, *singular value decomposition*) de la matriz  $\mathbf{X}$ . La SVD diagonaliza la matriz  $\mathbf{X}^\top \mathbf{X}$  y encuentra una descomposición de  $\mathbf{X}$  en tres matrices  $U, V$  y  $\Sigma$  tal que  $\mathbf{X} = U \Sigma V^\top$  donde  $V$  es la matriz de los vectores singulares ( $v_i$ ) puestos en columnas,  $\Sigma$  es un matriz en cuya diagonal los valores singulares ( $\sqrt{\lambda_i}$ ) ordenados de manera descendente y  $U$  esta construida de tal manera que sus columnas son  $X v_i / \sqrt{\lambda_i}$ .

El M-SSA aplica SVD a la matriz  $\mathbf{X}$  y resulta que  $\mathbf{X} = U \Sigma V^\top / \sqrt{\eta}$ , donde  $\eta$  es un factor de normalización que es igual al valor más grande entre  $N$  y  $D \times M$ . Los autovectores  $v_k$  forman una nueva base ortogonal en el espacio de embebido por retardo temporal de  $\mathbf{X}$ . Los autovalores  $\lambda_k$  se pueden interpretar como la varianza en la dirección de  $v_k$ .

Proyectando las series temporales  $\mathbf{X}$  en los autovectores,

$$\mathbf{A} = \mathbf{X} \mathbf{V}$$

se obtienen los componentes principales (PCs, principal componentes) como las columnas de  $\mathbf{A}$ . Los PCs tienen el mismo numero de filas que  $\mathbf{X}$ , y no tienen correlación a retardo

nulo; pueden ser considerados como componentes filtradas de las series temporales  $x$ , con filtros adaptados a los datos dados por los autovectores. Esta propiedad de filtrado se vuelve evidente si re-escribimos la proyección de los autovectores explícitamente como:

$$a_k(n) = \sum_{d=1}^D \sum_{m=1}^M x_d(n+m-1)v_{dk}(m), \quad (1.4)$$

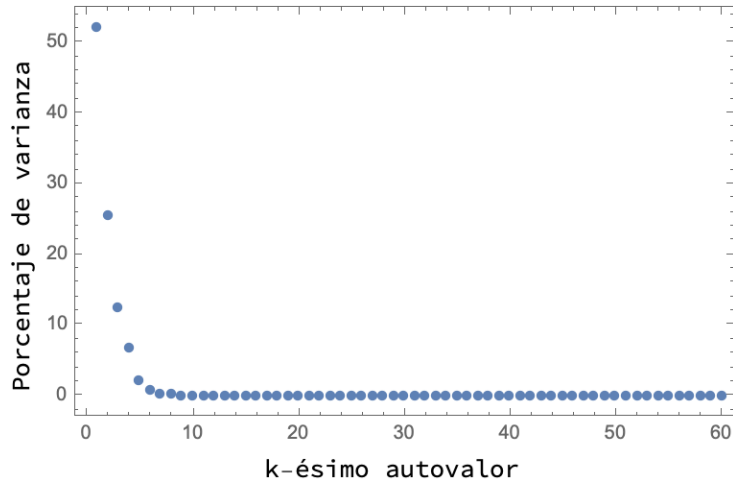
con  $k = 1, \dots, DM$  y  $n = 1, \dots, N - M + 1$ . En particular, con  $M = 1$  se reduce el M-SSA a un análisis de componentes principales (*Principal Component Analysis*, PCA en inglés) clásico para  $D$  variables.

Los coeficientes  $a_k(\cdot)$  son las entradas de la matriz de componentes principales  $Y$ . La notación  $V_k = \{v_{dk}(m) : 1 \leq d \leq D\}$  permite reflejar la naturaleza especial de los autovectores, que están compuestos de  $D$  segmentos consecutivos de longitud  $M$ , cada uno de los cuales está asociado a un canal,  $v_{dk}(m) \equiv v_k(\{d-1\}M + m)$ . De este modo, la sumatoria sobre  $m$  en la ecuación (1.4) puede ser interpretada como una respuesta clásica a un filtro de impulso finito (FIR en inglés, finite-impulse response) que opera en el canal  $d$  con coeficientes de filtrado dados por los elementos del autovector  $v_{dk}(m)$ . La segunda sumatoria, sobre  $d$ , que combina los canales filtrados, representa un PCA clásico.

Una forma de reconstruir los componentes individuales del comportamiento del sistema que es óptima en el sentido de mínimos cuadrados viene dada por la transformación:

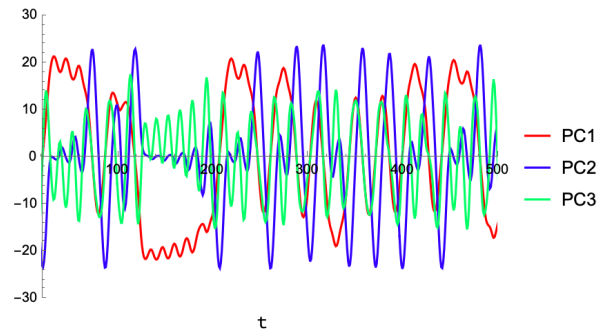
$$r_{dk}(n) = \frac{1}{M_n} \sum_{m=L_n}^{U_n} a_k(n-m+1)e_{dk}(m). \quad (1.5)$$

A los  $r_{dk}(n)$  se los llama componentes reconstruidos (RCs, *reconstructed components* en inglés), y representan la parte del canal  $x_d$  que corresponde al par  $(\lambda_k, e_k)$ . Los valores del factor de normalización  $M_n$  y de los límites de las sumatoria  $L_n$  y  $U_n$  para la parte central de la serie temporal,  $M \leq n \leq N - M + 1$ , son simplemente  $(M_n, L_n, U_n) = (M, 1, M)$ ; dados en (Ghil et al., 2002) para ambos extremos. En particular, la serie temporal puede reconstruirse completamente mediante la suma de todas sus RCs.

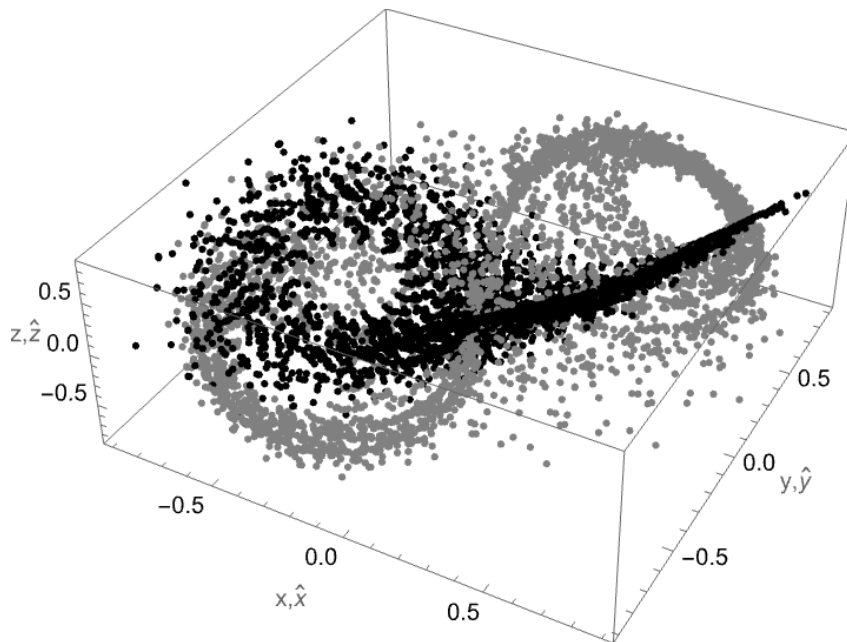


**Figura 1.5:** Porcentajes de varianza asociados a los autovalores del espectro multicanal M-SSA del atractor de Lorenz para una nube tridimensional de 3901 puntos.

Para ilustrar la técnica, la Figura 1.5 muestra el porcentaje de varianza asociado a cada valor singular tomando como *input* las tres series temporales  $x(t)$ ,  $y(t)$ ,  $z(t)$  del atractor de



**Figura 1.6:** Primeros tres componentes principales extraídos del M-SSA en el caso del atractor de Lorenz.



**Figura 1.7:** Nube de puntos reconstruida ( $\hat{x}, \hat{y}, \hat{z}$ , en gris) por M-SSA usando los primeros 6 RCs, sobre la nube de puntos original ( $x, y, z$  en negro) del atractor de Lorenz.

Lorenz determinístico. La Figura 1.6 muestra los tres primeros componentes principales. La Figura 1.7 muestra dos nubes de puntos: en negro, la nube de puntos original, y en gris: un embebido por retardos de la reconstrucción utilizando los primeros seis RCs en un M-SSA con tres canales (uno para  $x$ , uno para  $y$  y otro para  $z$ ). La nube reconstruida no coincide exactamente con la original – se observa la estructura de la mariposa deformada y rotada, pero la topología de la variedad enramada que subyace a ambas nubes de puntos no se altera.

Las series temporales de los atractores climáticos que deberemos analizar tienen una complejidad mucho mayor que las de un atractor caótico generado con tres ecuaciones diferenciales ordinarias como el de Lorenz. Es en estos casos en los que la utilidad del M-SSA se pone especialmente de manifiesto, puesto que permite obtener –al retener solamente algunas componentes reconstruidas– series temporales simplificadas respecto de las observadas, operando una reducción de la dimensionalidad del sistema y simplificando en consecuencia la variedad enramada subyacente. Es preciso tener presente que el análisis topológico realizado sobre datos pre-procesados con M-SSA proveerá una descripción de los mecanismos que fueron retenidos junto con las componentes reconstruidas seleccionadas.



## Capítulo 2

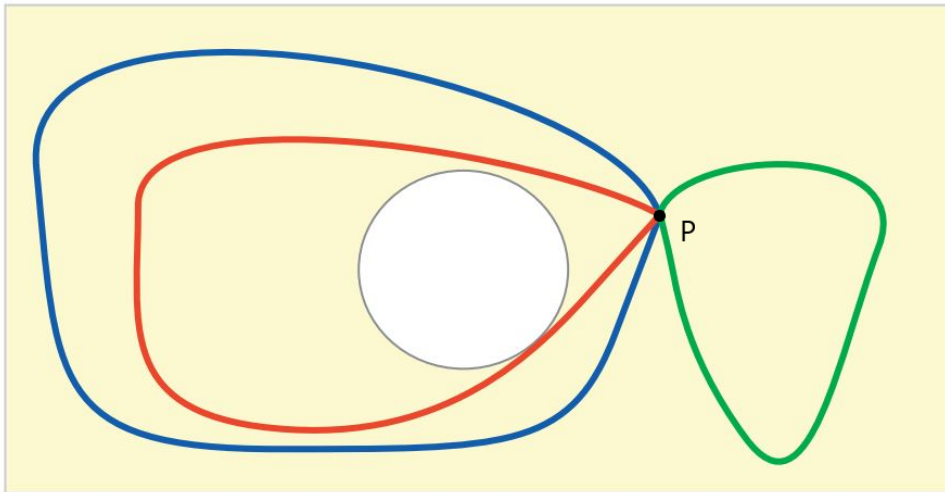
# Homologías persistentes

El uso de la topología algebraica ha tenido un éxito sorprendente en el análisis de una amplia variedad de datos no estructurados. Reunidos en un campo que hoy se denomina Análisis Topológico de Datos (Topological Data Analysis, TDA), proliferan los métodos que permiten estudiar grandes conjuntos de datos enfocándose en la conectividad de dicho conjunto (Chazal and Michel, 2021; Wasserman, 2018; Murugan and Robertson, 2019). Con el objetivo de detectar patrones de cualquier índole en una nube de puntos, estos métodos se utilizan cada vez más en numerosos y diversos ámbitos de investigación, especialmente en todo lo relacionado con el procesamiento de imágenes (Carlsson and Zomorodian, 2007).

Aunque TDA se refiere en general a todos los métodos que utilizan nociones de forma y conectividad para encontrar estructura en los datos (Wasserman, 2018), el método más popular se denomina homologías persistentes (Persistent Homologies, PH) (Edelsbrunner and Harer, 2008; Carlsson, 2014), al punto que TDA y PH son a veces considerados sinónimos. En el campo particular del análisis de series temporales, existen algunos trabajos (Perea and Harer, 2015; Perea, 2019; Fernández et al., 2023) que muestran de qué modo es posible extraer información sobre patrones cuasi-periódicos y cambios en la dinámica de un sistema aplicando PH a nubes de puntos construidas utilizando embebidos.

Como hemos visto en el capítulo anterior, la topología de la variedad enramada que subyace un atractor aporta mucha información sobre la estructura global de un sistema dinámico. Cuando el punto de partida es una nube de puntos obtenida a partir de observaciones o medidas, la extracción de la variedad enramada subyacente con la construcción de un único complejo celular que la aproxime correctamente con unas pocas celdas, resulta muchas veces dificultoso, cuando no imposible. Esta dificultad se incrementa cuando las señales son ruidosas o cuando se trata de un sistema cuyo espacio de fases tiene una dimensión alta. En PH, la estrategia no es aproximar variedades enramadas subyacentes, sino construir familias de complejos y analizar rasgos persistentes a través de las escalas. Estos rasgos persistentes no permiten, al menos de un modo trivial o inmediato, obtener un complejo celular representativo de la variedad enramada, como ocurre con el complejo BraMAH de la Figura 1.4 para las series temporales del atractor de Lorenz. No obstante, estos rasgos ofrecen una caracterización topológica que puede calcularse para cualquier nube, sin mediar ninguna hipótesis previa sobre la distribución local de los puntos. En vistas de la posibilidad de utilizar esta información para estimar las propiedades que debería tener la variedad enramada subyacente, cobra sentido investigar si los atractores climáticos pueden distinguirse en términos de sus homologías persistentes.

En esta tesis utilizaremos entonces PH para obtener marcadores topológicos que nos permitan lograr una primera descripción de los rasgos persistentes para nubes de puntos



**Figura 2.1:** Tres bucles (rojo, azul y verde) que empiezan y terminan en  $P$  en el espacio  $\hat{R}$ . Los bucles rojo azul son equivalentes porque es posible deformar el azul en el rojo sin que ninguno salga del espacio  $\hat{R}$ . Por otro lado, los bucles rojo y verde se consideran no equivalentes porque para deformar el rojo en el verde se debe salir del espacio  $\hat{R}$ . Fuente: Carlsson (2009).

asociadas a embebidos obtenidos a partir de series temporales simplificadas para los diferentes atractores climáticos. En este capítulo se presentan los elementos fundamentales de la teoría de homología persistente. La primera sección expone las nociones elementales de homotopía, homología y complejo simplicial. En la segunda sección, encontrarán los conceptos y las proposiciones de homología persistente, necesarias para describir y comparar las propiedades topológicas de las nubes de puntos a estudiar en esta tesis.

## 2.1. Topología algebraica

La topología algebraica nos proporciona, entre otras cosas, la matemática que formaliza la idea de conectividad en un espacio topológico. Esta información de conectividad se aborda cuando se analizan los agujeros del espacio topológico: las componentes arcoconexas (los agujeros de dimensión 0), los ciclos (los agujeros de dimensión 1), las cavidades (los agujeros de dimensión 2) y las hipercavidades (los agujeros de dimensión  $n$ ). Con el fin de distinguir o encontrar similitudes entre los espacios, la topología algebraica utiliza métodos para contar las ocurrencias de patrones geométricos en un espacio. Un invariante topológico muy utilizado en la topología algebraica para clasificar espacios topológicos son los grupos de homotopía que registran información sobre los agujeros de un espacio topológico.

### 2.1.1. Homotopías

El primer grupo de homotopía es el más conocido y se denomina grupo fundamental. A modo de ejemplo, consideremos el espacio  $\hat{R}$  que resulta de quitar un disco circular al rectángulo sombreado en amarillo, el punto  $P$  y tres bucles (rojo, azul y verde) que empiezan y terminan en  $P$  como se muestra en la Figura 2.1. Dos bucles pueden combinarse recorriendo el primero y luego el segundo y se consideran equivalentes si uno puede deformarse en el otro sin romperse. Los bucles rojo azul son equivalentes porque es posible deformar el azul en el rojo sin que ninguno salga del espacio  $\hat{R}$ . Por otro lado, los bucles rojo y verde se consideran no equivalentes porque para deformar el rojo en el verde se debe salir del espacio  $\hat{R}$ . El conjunto de todos los bucles que empiezan y terminan en  $P$ , el método de combinación entre los bucles y la relación de equivalencia es el grupo fundamental para este espacio. El recuento de bucles en un espacio topológico, es una cantidad interesante a



considerar para la distinción entre espacios.

La formalización de un bucle en un espacio  $X$  es un ciclo, que podemos entender como a un mapa continuo  $f : S_1 \rightarrow X$ , donde  $S_1$  es la circunferencia. En la Figura 2.1 los ciclos azul y rojo permiten capturar un rasgo característico del espacio  $\hat{R}$  que es el agujero en el centro. Al igual que los ciclos azul y rojo, existen potencialmente infinidad de otros ciclos que “rodean” a este agujero y que permiten percibirlo. La idea es capturar este agujero utilizando la presencia de ciclos a su alrededor, pero como contar ocurrencias de patrones directamente es inviable, una buena estrategia para distinguir espacios es contar las clases de equivalencia de ocurrencias de patrones bajo una relación de equivalencia, como lo es la relación de ser homotópico. El conjunto de clases de equivalencia se denomina grupo de homotopía.

**Definición 6.** Sea  $f, g : X \rightarrow Y$ , dos mapas continuos, se dice que  $f$  y  $g$  son homotópicas, y escribimos  $f \simeq g$ , si existe un mapa continuo  $H : X \times [0, 1] \rightarrow Y$  tal que  $H(x, 0) = f(x)$  y  $H(x, 1) = g(x)$ .  $H$  recibe el nombre de homotopía. La relación de ser homotópico es una relación de equivalencia.

**Definición 7.** Sea  $f : X \rightarrow Y$ , un mapa continuo, se dice que  $f$  es una equivalencia homotópica si existe un mapa continuo  $g : Y \rightarrow X$  tal que  $f \circ g$  es homotópica al mapa identidad en  $Y$  y  $g \circ f$  es homotópica a  $f$ .

**Definición 8.** Sean  $X$  e  $Y$  dos espacios y  $f : X \rightarrow Y$  una equivalencia homotópica, se dice que  $X$  e  $Y$  son homotópicamente equivalentes.

Un espacio que es homotópicamente equivalente a un punto se dice que es un espacio contráctil, es decir, podemos reducir a  $X$  a un punto mediante una homotopía. Por ejemplo, el plano y la esfera son espacios contráctiles y tienen como grupo fundamental al trivial.

Los grupos de homotopía de dimensión mayor a 1, proporcionan invariantes que realizan la misma parametrización que el grupo fundamental, pero en el caso de ciclos de mayor dimensión, es decir, mapeos de esferas de mayor dimensión en el espacio. Estos grupos son habitualmente difíciles de calcular, sin embargo existen invariantes más computables, denominados grupos de homología, en los que la relación de equivalencia de los ciclos se extiende declarando que dos ciclos son equivalentes si existe una superficie con una frontera igual a la unión de los dos ciclos.

### 2.1.2. Homologías

Los grupos de homología son una serie de invariantes escalonados. Dichos invariantes codifican la información sobre la cantidad de componentes conexas que tiene la variedad estudiada y sobre la existencia y número de agujeros de distintas dimensiones. Inicialmente, la homología se definió para espacios descritos como complejos simpliciales. Sin embargo, no todos los espacios pueden describirse como complejos simpliciales, y además, los espacios pueden describirse como complejos simpliciales de muchas maneras distintas. Estos problemas se resuelven extendiendo la definición de grupos de homología a todos los espacios. De esta forma, el resultado ya no depende de ninguna estructura particular como un complejo simplicial, sino sólo del espacio en sí. Dado que la mayoría de los espacios de interés son complejos simpliciales explícitos u homotópicamente equivalentes a dichos complejos, resulta que el cálculo simplicial es suficiente para la mayoría de las situaciones.

## Complejos simpliciales

Desde un punto de vista geométrico, un complejo simplicial es un espacio topológico construido con “ bloques ” de distintas dimensiones llamados símlices. Estos símlices pueden ser vértices, aristas, triángulos, tetraedros y análogos de dimensiones mayores, que se van pegando a lo largo de las caras para construir así al complejo. Si dos símlices se intersecan, la intersección es una cara común. En la Figura 2.2 vemos complejos simpliciales, uno válido y otro no válido, donde la intersección no es una cara común. En lo que sigue vamos a formalizar la definición de complejo simplicial.

Sea  $S = \{x_0, x_1, \dots, x_n\} \subseteq \mathbb{R}^k$ , decimos que  $S$  está en *posición general* si no está contenido en ningún hiperplano afín de  $\mathbb{R}^k$  de dimensión menor que  $n$ . Cuando  $S$  está en posición general, definimos como *símplice generado por  $S$*  a la cápsula convexa  $\sigma = \sigma(S)$  de  $S$  en  $\mathbb{R}^k$ . Los puntos  $x_i$  se denominan *vértices*, y los símlices  $\sigma(T)$  generados por subconjuntos no vacíos de  $T \subseteq S$  se denominan *caras* de  $\sigma$ .

### Definición 9. Complejo simplicial

Un complejo simplicial (finito) es una colección finita  $\chi$  de símlices que cumplen:

- (i) Para cualquier símlice  $\sigma$  de  $\chi$ , todas las caras de  $\sigma$  están también contenidas en  $\chi$ .
- (ii) Para cualesquiera dos símlices  $\sigma$  y  $\tau$  de  $\chi$ , la intersección  $\sigma \cap \tau$  es un símlice que es una cara tanto de  $\sigma$  como de  $\tau$ .

Es un caso particular de la definición 5 donde las celdas son símlices.

Todo complejo simplicial determina un objeto combinatorio formado por subconjuntos del conjunto de vértices del complejo. Esta observación nos lleva a una definición abstracta de complejo simplicial.

### Definición 10. Complejo simplicial abstracto

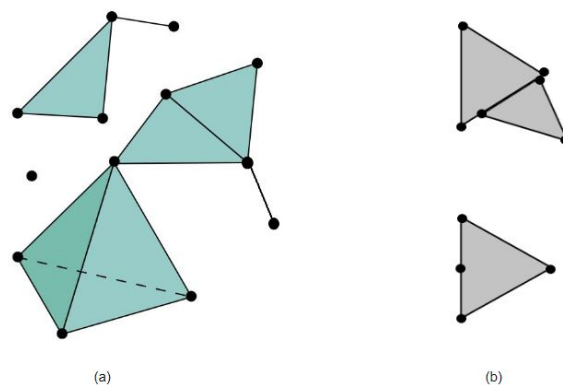
Un complejo simplicial abstracto (finito) es un par  $(V, \Sigma)$ , donde  $V$  es un conjunto finito y  $\Sigma$  es una familia de subconjuntos no vacíos de  $V$  tales que si  $\sigma \in \Sigma$  y  $\tau \subseteq \sigma$  entonces  $\tau \in \Sigma$ .

Un complejo simplicial  $\chi$  determina un complejo simplicial abstracto. Los vértices  $V$  serán todos los vértices de todos los símlices de  $\chi$  y un subconjunto de  $V$  es un símlice de  $\Sigma$  si y solo si es un conjunto de vértices de algún símlice de  $\chi$ . También es cierto que un complejo simplicial abstracto determina el espacio subyacente del complejo simplicial hasta un homeomorfismo. Dado un complejo simplicial abstracto  $X$ , le asociamos un espacio  $|X|$  al que llamamos la realización geométrica de  $X$ , y todo complejo simplicial es homeomorfo a la realización geométrica de su complejo simplicial abstracto asociado (para más detalles remitimos al lector a (Carlsson, 2014) y (Munkres, 2018)).

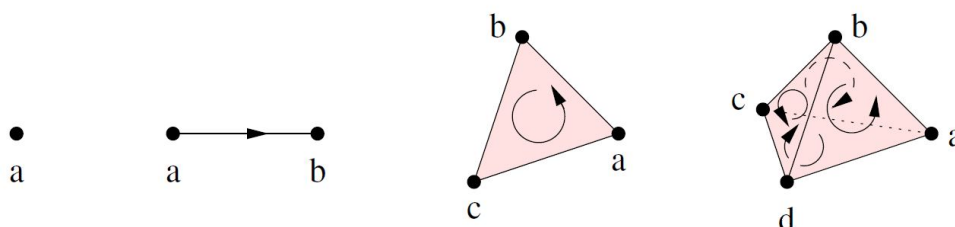
Los complejos simpliciales constituyen una forma combinatoria sencilla de describir determinados espacios topológicos. Una ventaja de utilizar estos complejos, es que la homología puede calcularse utilizando sólo el álgebra lineal de los módulos  $\mathbb{Z}$  finitamente generados.

Dado cualquier complejo simplicial  $X = (V, \Sigma)$ , llamamos *k-símlices* de dimensión  $k$  a los  $\sigma \in \Sigma$  tales que la cantidad de caras de  $\sigma$  es  $k + 1$ . Llamamos  $\Sigma_k$  al subconjunto de  $\Sigma$  conformado por todos los *k-símlices*. Como ejemplos tenemos: un *vértice* es un 0-símlice, una *arista* es un 1-símlice, un *triángulo* es un 2-símlice y un *tetraedro sólido* es un 3-símlice. Vemos estos ejemplos en la Figura 2.3. La dimensión de un complejo se define como el máximo de las dimensiones de todos los símlices que lo componen.

En la definición de complejo simplicial abstracto, un símlice está determinado por un conjunto de vértices donde el orden es irrelevante,  $\{x_1, x_2\}$  y  $\{x_2, x_1\}$  definen el mismo



**Figura 2.2:** En un complejo simplicial válido dos símlices se intersecan en una cara común. En (a) vemos un complejo simplicial válido, las intersecciones son caras comunes. En (b) tenemos un complejo donde los símlices que lo forman comparten parcialmente sus caras. Decimos que es un complejo simplicial no válido.



**Figura 2.3:** Ejemplos de  $k$ -símlices orientados en  $\mathbb{R}^3$ , para  $0 \leq k \leq 3$ . Un vértice  $a$  es un 0-símplex, un segmento  $[a, b]$  es un 1-símplex, un triángulo  $[a, b, c]$  es un 2-símplex y un tetraedro sólido  $[a, b, c, d]$  es un 3-símplex. A cada símplex se le asigna una orientación denotada por una flecha. Fuente: Zomorodian and Carlsson (2005).

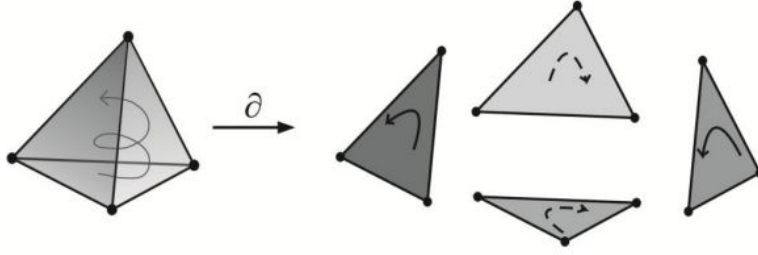
símplex. Para definir homología necesitamos que los símlices estén ordenados, es decir necesitamos darle una orientación a todas las celdas del complejo.

### Grupos de homología

Sea  $\tau : \{0, 1, \dots, n\} \rightarrow \{0, 1, \dots, n\}$  una permutación, una *orientación* de un  $k$ -símplex  $\sigma$ ,  $\sigma = \{x_0, \dots, x_n\}$ , es una clase de equivalencia de ordenaciones de los vértices de  $\sigma$ , donde  $(x_0, \dots, x_n) \sim (x_{\tau(0)}, \dots, x_{\tau(n)})$  son equivalentes si y sólo si  $\tau$  es par. Recordemos que una permutación es par si se descompone como el producto de una cantidad par de transposiciones. Dos ordenamientos de vértices se dicen equivalentes si difieren en una permutación par. Los ordenamientos de los vértices de  $\sigma$  caen entonces en dos clases de equivalencia, excepto si  $\sigma$  es un 0-símplex, entonces hay una sola clase y por lo tanto una sola orientación de  $\sigma$ . Un símplex  $\sigma$  junto con una orientación elegida se denomina *símplex orientado* y se denotará por  $[\sigma]$ . En la Figura 2.3 vemos una posible orientación de los símlices.

Definimos el grupo de  $k$ -cadenas en  $X$ ,  $C_k(X)$ , como el grupo de combinaciones lineales formales de elementos en  $\Sigma_k$  con coeficientes en  $\mathbb{Z}$ , o equivalentemente el grupo abeliano libre sobre el conjunto  $\Sigma_k$ . Un elemento  $c \in C_k(X)$  es una  $k$ -cadena tal que  $c = \sum_i n_i [\sigma_i]$ ,  $\sigma_i \in \Sigma_k$  con coeficientes  $n_i \in \mathbb{Z}$ .

Si imponemos un orden total en el conjunto de vértices  $V$ , definimos el operador  $d_i : \Sigma_k \rightarrow \Sigma_{k-1}$ , para  $0 \leq i \leq k$ , siendo  $d_i(\sigma) = \sigma - \{s_i\}$ , donde  $s_i$  denota el  $i$ -ésimo elemento en  $\sigma$ , bajo la ordenación total dada. Definimos el operador lineal *borde*, también llamado



**Figura 2.4:** El operador borde  $\partial_k$  asigna un 3-símplice a cuatro 2-símplices. Fuente: Maletić et al. (2016).

frontera, como  $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$  como

$$\partial_k(\sigma) = \sum_{i=0}^k (-1)^i d_i(\sigma).$$

Dada una  $k$ -cadena  $c$ , el elemento  $\partial_k(c)$  se llama borde de  $c$ . Una  $k$ -cadena con borde nulo se llama  $k$ -ciclo. En la Figura 2.4 se observa cómo el operador borde asigna un 3-símplice a cuatro 2-símplices.

**Ejemplo 1.** Para un 1-símplice tenemos  $\partial_1([x_0, x_1]) = x_1 - x_0$ . Para un 2-símplice tenemos  $\partial_2([x_0, x_1, x_2]) = [x_0, x_1] + [x_1, x_2] - [x_0, x_2]$

El operador borde conecta los grupos de cadenas en un complejo de cadenas  $C_*$ :

$$\dots \rightarrow C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \rightarrow \dots$$

Una característica de los complejos de cadenas es la siguiente:

**Lema 1.**  $\partial_k \circ \partial_{k+1} \equiv 0$ . Se verifica de manera inmediata por la definición de borde dada más arriba.

Podemos definir subgrupos de  $C_k$  utilizando el operador borde. El núcleo de  $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$  se llama el grupo de  $k$ -ciclos y se denota por  $Z_k(X) = \text{Núcleo}(\partial_k)$ . La imagen de  $\partial_{k+1} : C_{k+1}(X) \rightarrow C_k(X)$  se llama el grupo de  $k$ -bordes y se denota por  $B_k(X) = \text{Imagen}(\partial_{k+1})$ . Es decir, en  $B_k(X)$  se encuentran los  $k$ -ciclos que son borde de una  $k + 1$ -cadena.

Del lema 1 se deduce que  $\text{Imagen}(\partial_{k+1}) \subseteq \text{Núcleo}(\partial_k)$ . Es decir, cada borde de cualquier  $k + 1$ -cadena es automáticamente un  $k$ -ciclo. En otras palabras, los subgrupos están anidados,  $B_k \subseteq Z_k \subseteq C_k$ .

### Definición 11. Grupo de Homología

Se define el  $k$ -ésimo grupo de homología como:

$$H_k(X) \cong \text{Núcleo}(\partial_k) / \text{Imagen}(\partial_{k+1}) = Z_k(X) / B_k(X)$$

Recordemos que  $\text{Núcleo}(\partial_k)$  es el grupo de todos los  $k$ -ciclos, mientras que  $\text{Imagen}(\partial_{k+1})$  es el grupo de los  $k$ -ciclos que son borde de alguna  $k + 1$ -cadena. Por lo tanto, el cociente  $\text{Núcleo}(\partial_k) / \text{Imagen}(\partial_{k+1})$  sólo conserva los ciclos que no son borde de ninguna cadena de dimensión mayor. En otras palabras, se conservan sólo los ciclos que representan “agujeros”. Los grupos de homología reúnen toda la información esencial que se tiene sobre los

complejos: las componentes conexas (agujeros de dimensión 0 ó 0-agujeros), los ciclos (agujeros de dimensión 1 ó 1-agujeros), las cavidades que se forman (agujeros de dimensión 2 ó 2-agujeros) y las hipercavidades (agujeros de dimensión  $n$  ó  $n$ -agujeros).

Ahora podemos definir un invariante topológico llamado número de Betti, que nos permitirá distinguir espacios topológicos.

## Números de Betti

La dimensión de  $H_k$ , es decir, el número de generadores del  $k$ -ésimo grupo de homología, se denomina el  $k$ -ésimo número de Betti:

$$\beta_k(X) = \dim(H_k) \tag{2.1}$$

El número de Betti  $\beta_0$  nos da el número de componentes conexas de  $X$ . Si el complejo está formado por una única componente,  $\beta_0$  valdrá 1, si está compuesto por  $m$  componentes,  $\beta_0$  valdrá  $m$ . En el caso de  $\beta_1$ , tendremos los 1-agujeros y con  $\beta_2$  los 2-agujeros.

A continuación vemos los números de Betti para superficies conocidas. Observar que complejos diferentes, por ejemplo el circunferencia y el cilindro pueden tener los mismos grupos de homología.

**Cuadro 2.1:** Números de Betti  $\beta_k$  para algunas superficies.

	$\beta_0$	$\beta_1$	$\beta_2$
Disco	1	0	0
Esfera	1	0	1
Circunferencia	1	1	0
Cilindro	1	1	0
Banda de Möbius	1	1	0
Toro	1	2	1

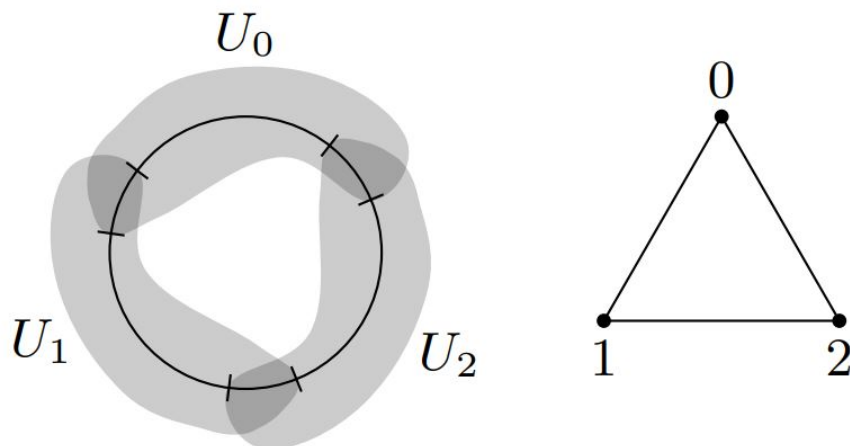
## 2.2. Nervio de un cubrimiento

Dado que la homología de los complejos simpliciales es algorítmicamente computable, se desea entonces construir complejos simpliciales que computen la homología de un espacio subyacente  $X$  o al menos tengan una fuerte relación con él. Hay una serie de complejos simpliciales que pueden construirse a partir de  $X$ . Para definir estos complejos, primero definiremos el nervio de un cubrimiento de un espacio topológico  $X$ , que nos permitirá asociar un espacio topológico a un complejo simplicial a través de un cubrimiento.

### Definición 12. Nervio de un cubrimiento

Sea  $X$  un espacio topológico y sea  $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$  un cubrimiento de  $X$ . Denominaremos  $\mathcal{N}(\mathcal{U})$ , nervio de  $\mathcal{U}$ , al complejo simplicial abstracto con conjunto de vértices  $A$ , y donde una familia  $\{\alpha_0, \dots, \alpha_k\}$  genera un  $k$ -símplice si y sólo si  $U_{\alpha_0} \cap \dots \cap U_{\alpha_k} \neq \emptyset$ .

La definición general de nervio requiere trabajar con complejos infinitos. Sin embargo, en la práctica, se suele trabajar con variedades compactas donde siempre se puede tomar cubrimientos finitos. El “teorema del nervio” proporciona criterios que garantizan que  $\mathcal{N}(\mathcal{U})$



**Figura 2.5:** A la izquierda de la figura se puede observar una circunferencia y un cubrimiento  $\mathcal{U}$  que cumple con las propiedades del teorema del nervio. A la derecha se puede ver el complejo simplicial asociado al nervio del cubrimiento  $\mathcal{U}$ . El complejo que se obtiene es homotópicamente equivalente a la circunferencia. Fuente: Díaz Sánchez et al. (2019).

es homotópicamente equivalente al espacio subyacente  $X$ . En la Figura 2.5 se puede observar un ejemplo de un cubrimiento de una circunferencia. El complejo simplicial asociado al nervio del cubrimiento, es un triángulo cuyos tres vértices corresponden a las intersecciones del nervio. Este complejo es homotópicamente equivalente a la circunferencia.

**Teorema 1.** Sean  $X$  un espacio topológico y  $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$  un cubrimiento de  $X$ , y supongamos que el cubrimiento consiste en conjuntos abiertos y es numerable. Supongamos además que para todo  $\emptyset \neq S \subseteq A$ , tenemos que  $\bigcap_{s \in S} U_s$  es o bien contráctil o bien vacío. Entonces,  $\mathcal{N}(\mathcal{U})$  es homotópicamente equivalente a  $X$ .

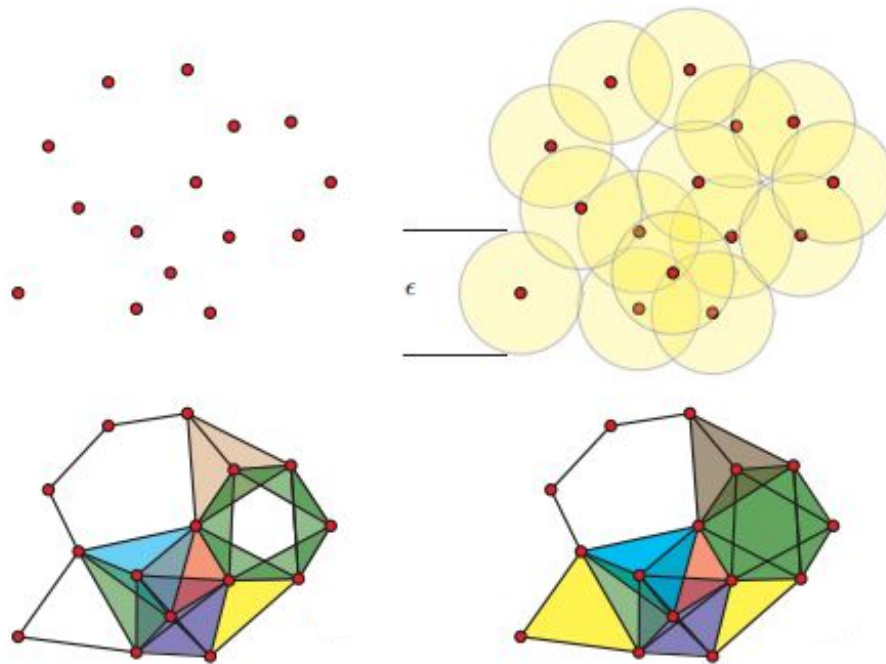
En la próxima sección definiremos dos de los cubrimientos más utilizados en espacio métricos: el complejo de Čech y el complejo de Vietoris-Rips.

### 2.3. Complejos Čech y Vietoris-Rips

Cuando el espacio en cuestión es un espacio métrico, un cubrimiento viene dado por la familia  $\mathcal{B}_\epsilon(X) = \{B_\epsilon(x)\}_{x \in X}$ , para algún  $\epsilon > 0$ . Más en general, podemos tomar cualquier subconjunto  $V \subseteq X$  en el que  $X = \bigcup_{v \in V} B_\epsilon(v)$  y construir el nervio del cubrimiento  $\{B_\epsilon(v)\}_{v \in V}$ . Denotaremos esta construcción por  $\check{C}(V, \epsilon)$ , y nos referiremos a ella como el complejo de Čech asociado a  $V$  y  $\epsilon$ . Este complejo es el complejo simplicial más elemental que se puede construir sobre  $X$ .

**Teorema 2.** Sea  $M$  una variedad de Riemann compacta. Entonces existe un número positivo  $e$  tal que  $\forall \epsilon \leq e$ ,  $\check{C}(M, \epsilon)$  es homotópicamente equivalente a  $M$ . Además, para cada  $\epsilon \leq e$ , existe un subconjunto finito  $V \subseteq M$  tal que el subcomplejo de  $\check{C}(V, \epsilon) \subseteq \check{C}(M, \epsilon)$  también es homotópicamente equivalente a  $M$ .

Es decir, que bajo la hipótesis de que los puntos de  $V$  están bien distribuidos podemos conseguir un complejo homotópicamente equivalente a  $M$ . Uno de los problemas de esta construcción es que es costosa desde el punto de vista computacional, ya que requiere el almacenamiento de símlices de varias dimensiones. Una idea para lidiar con ese problema es construir un complejo simplicial que pueda recuperarse únicamente a partir de la información de las aristas. Esto sugiere la siguiente variante de la construcción de Čech denominada complejo de Vietoris-Rips.



**Figura 2.6:** Dada una nube de puntos, se construye el complejo de Čech (abajo a izquierda) y el complejo Vietoris Rips (abajo a derecha) donde el radio de la bola es  $\epsilon/2$ . Fuente: Ghrist (2008).

**Definición 13.** Sea  $X$  un espacio métrico, con métrica  $d$ . Entonces el complejo de Vietoris-Rips para  $X$ , asociado al parámetro  $\epsilon$  ( $\epsilon$  número real positivo), denotado por  $VR(X, \epsilon)$ , será el complejo simplicial cuyo conjunto de vértices es  $X$ , y donde cualquier subconjunto  $\{x_0, x_1, \dots, x_k\}$  de  $X$  genera un  $k$ -símplice si y sólo si  $d(x_i, x_j) \leq \epsilon$  para todo  $0 \leq i, j \leq k$ .

Observamos que los conjuntos de vértices de las dos construcciones son idénticos, por lo que ambos pueden considerarse subcomplejos del símplice  $X$ . En la Figura 2.6 se observan las construcciones de estos complejos sobre una nube de puntos. En ambas construcciones se consideran bolas con centro en cada punto de la nube. La figura de abajo a la izquierda muestra el complejo de Čech y la de abajo a la derecha el de Vietoris-Rips. El parámetro  $\epsilon$  para calcular el complejo Čech corresponde a los radios de las bolas, mientras que el parámetro del complejo Vietoris-Rips es la distancia entre los centros de las bolas. Se puede ver que los complejos obtenidos son diferentes y que el complejo de Čech está contenido en el de Vietoris-Rips.

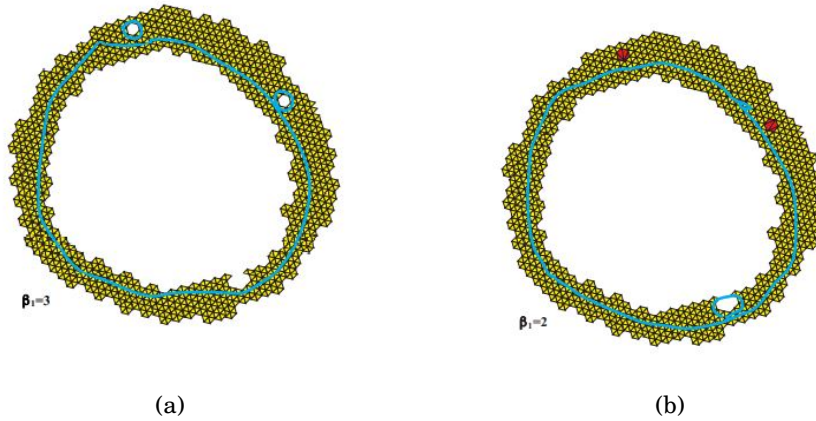
**Proposición 1.**  $\check{C}(X, \epsilon) \subseteq VR(X, 2\epsilon) \subseteq \check{C}(X, 2\epsilon)$ .

Para formar un símplice en el complejo de Čech basta con que las bolas tengan intersección, mientras que para los del complejo de Vietoris-Rips se necesita que los vértices del símplice estén dentro de dicha intersección (tomando las bolas cerradas). El complejo de Vietoris-Rips sigue siendo costoso desde el punto de vista computacional, debido a que su conjunto de vértices consiste en todo el espacio métrico en cuestión. Este complejo es una aproximación del complejo de Čech que no conserva necesariamente sus propiedades topológicas exactas, pero al ser sencilla su construcción, es uno de los más usados.

## 2.4. Homologías persistentes

Las homologías persistentes se utilizan en el Análisis Topológico de Datos (TDA) como método para estudiar los rasgos topológicos de los complejos celulares, observando cuáles



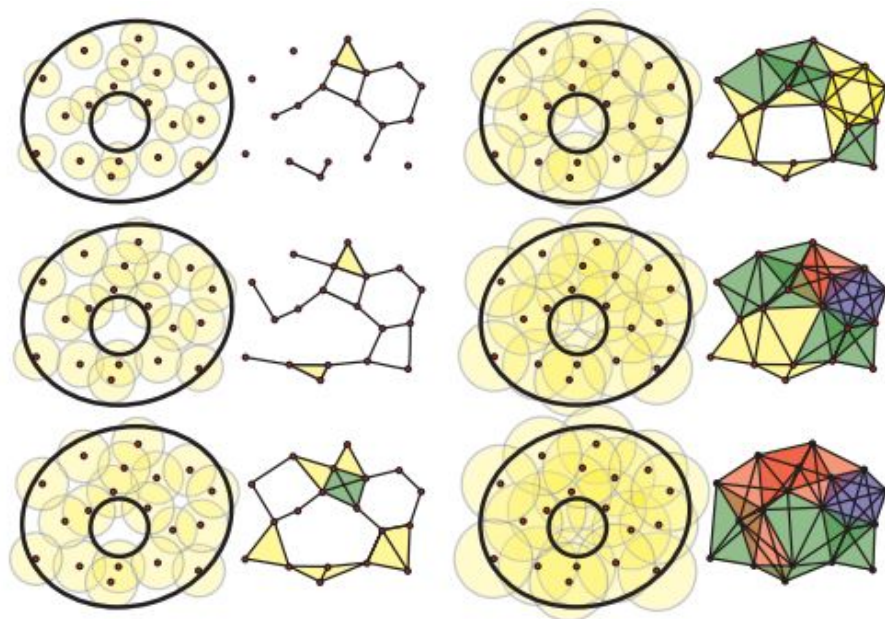


**Figura 2.7:** Complejos de Čech sobre una colección finita de puntos en el plano para dos valores diferentes de parámetro  $\epsilon$ . En el complejo (a) tenemos el primer número de Betti con valor tres y en el complejo (b) el primer número de Betti toma el valor dos. Fuente: Carlsson (2009).

de esos rasgos perduran durante la construcción de un complejo a medida que se varía cierto parámetro de filtración. Si tomamos los rasgos topológicos del complejo de corta y larga duración, podemos suponer que los de larga duración coinciden con las variables intrínsecas del sistema subyacente (los grupos de homología) mientras que los rasgos de corta duración se filtran. La construcción de un complejo simplicial es intrínsecamente arbitraria (Maletić et al., 2016). Un método para lograr este complejo es construirlo de forma gradual, comenzando por el más simple hasta llegar al último, que es estable estructuralmente. De esta forma se obtiene una secuencia de subcomplejos anidados. Cada subcomplejo está contenido en el siguiente creando una sucesión creciente de complejos. Esta sucesión se puede interpretar como una adición de nuevos símlices que se obtienen cambiando el parámetro de filtración. En la transición entre una etapa de filtración y la siguiente, ocurren cambios que son topológicamente relevantes como el nacimiento o muerte de  $n$ -agujeros  $\forall n \in \mathbb{N}_0$ . Estos cambios generan el nacimiento de nuevas clases de homología y la muerte de otras. Este método es el núcleo del enfoque de la topología algebraica computacional, y es lo que llamamos homología persistente. Existe una forma convencional para visualizar los intervalos de persistencia de los generadores de los grupos de homología. Estos intervalos se inician con el valor del parámetro de filtración cuando nace el generador y finaliza con el valor del parámetro de filtración cuando muere. Esta visualización se denomina Código de Barras. La otra forma de representar la homología persistente es mediante los Diagrama de Persistencia. La idea es la misma pero usando puntos en la representación en lugar de segmentos.

Consideremos la Figura 2.7 (a), un complejo de Čech construido sobre una colección finita de puntos en el plano euclídeo. Observamos que el conjunto de puntos se concentra en torno a una circunferencia. Al calcular la homología de este complejo, tendremos el primer número de Betti  $\beta_1 = 3$ , que incluye el gran bucle principal y los dos bucles más pequeños correspondientes a los dos agujeros 1-dimensionales más pequeños del complejo. Intuitivamente, se puede pensar que estos dos agujeros 1-dimensionales proceden de un muestreo defectuoso o de otros errores en la recuperación de los datos. También se podría argumentar que se deben a una elección incorrecta del parámetro  $\epsilon$  y que debemos aumentar su valor. Al hacer esto, obtenemos la Figura 2.7 (b). Ahora los dos agujeros 1-dimensionales más pequeños se han “rellenado”, y se ha introducido un nuevo agujero 1-dimensional en la parte inferior derecha de la figura. Si calculamos la homología de este complejo, obtendremos un primer número de Betti  $\beta_1 = 2$ . Pero este resultado es incorrecto para ambos





**Figura 2.8:** Evolución de una filtración con complejos Vietoris-Rips sobre una nube de puntos que representa un anillo. Al aumentar  $\epsilon$ , los agujeros 1-dimensionales aparecen y desaparecen. Fuente: Ghrist (2008).

valores del parámetro, pues sabemos que una circunferencia tiene  $\beta_1 = 1$ . Sin embargo, observamos en la Figura 2.7, que hay una inclusión del complejo (a) en el complejo (b), ya que el primero corresponde a un valor de parámetro menor que el segundo. Por tanto, podemos preguntarnos por la imagen de la homología del complejo (a) en la homología del complejo (b). Los dos ciclos pequeños de (a) desaparecen en (b), ya que están rellenos. Por otra parte, el ciclo pequeño en el complejo (b) no está en la imagen de la homología del complejo (a), ya que la arista necesaria no se rellena en (a). Vemos entonces que la imagen consiste exactamente en el ciclo mayor, que es lo que consideramos la respuesta “correcta” en este caso (Carlsson, 2009).

### 2.4.1. Filtraciones

El procedimiento de formación gradual de un complejo simplicial se conoce con el nombre de filtración, del cual surgen una secuencia de subcomplejos anidados. Cada subcomplejo está contenido en el subcomplejo de la siguiente etapa de filtración. Más formalmente, sea  $X$  un complejo, una *filtración* es una familia de subcomplejos de  $X = \{X_i\}_{i \in \{0,1,\dots,n\}}$ , tales que

$$\emptyset = X_0 \subset X_1 \subset \dots \subset X_n = X$$

Cada subcomplejo  $X_i = X(\epsilon_i)$  depende del parámetro de filtrado  $\epsilon_i$ . Si  $\epsilon_i < \epsilon_j$  entonces  $X_i \subset X_j$ , pues podemos pensar que el subcomplejo  $X_j$  proviene de agregar celdas al complejo  $X_i$ . Algunos ejemplos de filtración son los complejos de Čech y Vietoris-Rips.

Para cada  $i \leq j$  se tiene que el mapeo de inclusión de los espacios  $X_i$  en  $X_j$  induce un mapeo entre  $H_k(X_i) \rightarrow H_k(X_j)$  para cada  $k$ . La filtración entonces corresponde a una secuencia de grupos de homología conectados por un mapeo:

$$0 = H_k(X_0) \rightarrow H_k(X_1) \rightarrow \dots \rightarrow H_k(X_n) = H_k(X)$$

Al pasar de  $X_{i-1}$  a  $X_i$  obtenemos nuevas clases de homología y perdemos otras que se vuelven triviales o se fusionan entre si. Otra información importante que se obtiene es el  $\epsilon_i$  en el cual la clase aparece y  $\epsilon_j$  en el cual la clase desaparece. En la Figura 2.8 vemos una filtración con complejos Vietoris-Rips sobre una nube de puntos que representa un anillo. Para  $\epsilon$  suficientemente pequeño, el complejo es un conjunto discreto. Al aumentar  $\epsilon$ , se conectan puntos, los agujeros aparecen y desaparecen. Si se busca eliminar los agujeros pequeños del interior del anillo, se puede hacer crecer  $\epsilon$  más aún, eventualmente se rellena el agujero grande que distingue el anillo del disco.

Los intervalos de Betti ayudan a describir cómo el grupo de homología  $X(\epsilon)$  cambia a medida que  $\epsilon$  aumenta. Un intervalo de Betti  $k$ -dimensional  $[\epsilon_0, \epsilon_f)$ , corresponde a un agujero  $k$ -dimensional que aparece cuando el parámetro de filtrado es  $\epsilon = \epsilon_0$ , permanece “abierto”  $\forall \epsilon_0 < \epsilon < \epsilon_f$  y se “cierra” cuando  $\epsilon = \epsilon_f$ , es decir desaparece. Estamos interesados en los intervalos de Betti que persisten para un intervalo de filtrado suficientemente grande. Se llama persistencia a la longitud del intervalo de Betti,  $pers(\gamma) = \epsilon_f - \epsilon_0$ . Si el agujero  $k$ -dimensional nace pero no muere, definimos a la persistencia como infinito. Remitimos al lector a Edelsbrunner and Harer (2022) para más detalles sobre homologías persistentes.

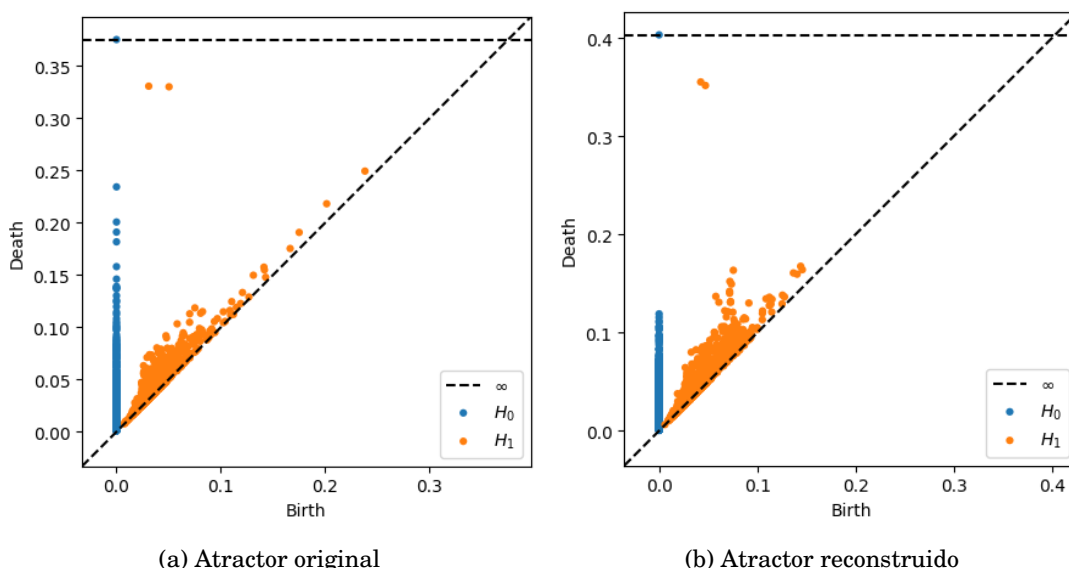
### 2.4.2. Diagramas de Persistencia

La homología persistente nos da entonces intervalos de persistencia de la forma  $[\epsilon_i, +\infty)$  y  $[\epsilon_i, \epsilon_j)$  que corresponden a los valores de filtración donde nacen las clases de homología y valores de filtración donde mueren las clases de homología. Para visualizar esta información contamos con los diagramas de persistencia.

El  $k$ -ésimo diagrama de persistencia de la filtración  $f$ , denotado como  $Dgmp_k(f)$ , se obtiene graficando cada punto  $(\epsilon_i, \epsilon_j)$  en el plano real extendido  $\overline{\mathbb{R}^2}$ , pues algunos de estos puntos pueden tener coordenada infinito. Cada punto en  $Dgmp_k(f)$  representa una clase  $\gamma$  y la distancia vertical a la diagonal es la persistencia. Notar que como  $i \leq j$ , todos los puntos se encuentran por encima de la diagonal. También es posible que existan otros generadores con punto asignado  $(\epsilon_i, \epsilon_j)$ , es por ello que  $(\epsilon_i, \epsilon_j)$  viene dado con una multiplicidad. Concretamente, la multiplicidad es el número de clases generadoras independientes que nacen en  $X_i$  y mueren entrando en  $X_j$ . Mientras mayor sea la persistencia, más tiempo persiste  $\gamma$  en la filtración. En el diagrama de persistencia los puntos cerca de la diagonal se consideran “ruido”.

En la Figura 2.9 se representan juntos a los diagramas de persistencia en dimensión  $k = 0$  (puntos azules) y en dimensión  $k = 1$  (puntos naranjas) del atractor de Lorenz y de su reconstrucción con M-SSA. Observemos primero el diagrama correspondiente al atractor original. A lo largo de la filtración se puede observar cómo van desapareciendo componentes conexas y uniéndose a otras hasta quedar una sola, representando un objeto conectado. Los agujeros 1-dimensionales aparecen en diferentes etapas, luego se van “rellenando”, y por tanto muriendo, pero sobreviven los dos agujeros característicos de la mariposa de Lorenz (Figura 1.2). En el atractor reconstruido, la distribución de los puntos naranjas y azules cambia un poco, puesto que la nube reconstruida está rotada y deformada respecto de la original, pero la existencia de dos agujeros 1-dimensionales de mayor persistencia se mantienen.

Como los puntos representados son de la forma  $(x, y)$  con  $x < y$ , la mitad triangular del diagrama de persistencia bajo la línea de referencia  $(y = x)$  siempre está sin utilizar y queda como espacio en blanco. Además los puntos de los diagramas de persistencia se solapan de forma natural para características similares y esto hace que algunos aspectos



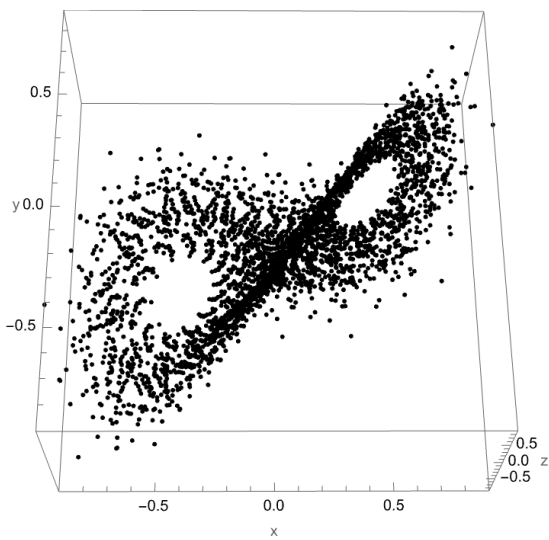
**Figura 2.9:** Diagramas de persistencia del atractor de Lorenz, donde cada punto representa el nacimiento y el tiempo de vida de la clase de homología. El diagrama (a) corresponde a la nube de puntos original y el diagrama (b) a la nube reconstruida. Ambas nubes se muestran –(a) en negro y (b) en gris– en la Figura 1.7. El punto azul superpuesto con la línea punteada representa al objeto conectado. Los dos puntos naranjas alejados de la diagonal representan los dos 1-agujeros de las alas de la mariposa de Lorenz, son las características que más persisten.

del diagrama sean poco claros. Sin embargo, el diagrama de persistencia puede visualizarse en forma plana como en la Figura 2.10. Este tipo de diagrama se llama diagrama de persistencia plana, cada punto es un par que representa el nacimiento y la persistencia de un generador de la clase de homología. Si  $\gamma$  nace en  $X_i$  y muere en  $X_j$ , el punto será  $(\epsilon_i, \epsilon_j - \epsilon_i)$ .

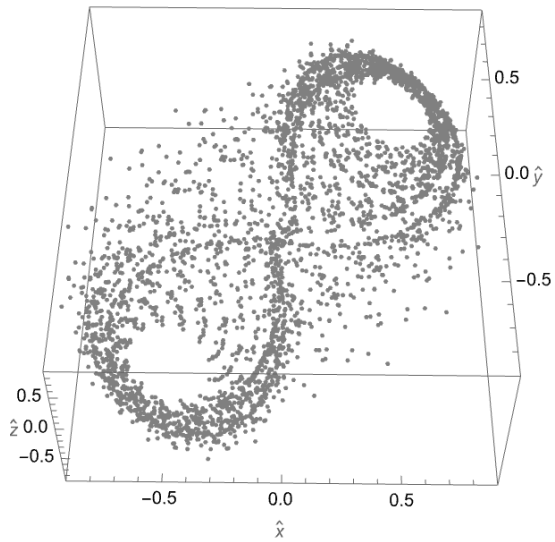
Los puntos cerca del eje horizontal se consideran “ruido”, ya que nacen y mueren rápido. Los puntos más alejados del eje horizontal representan características que persisten por más tiempo y resultan por tanto significativas. El diagrama de persistencia plana permite aprovechar mejor el espacio no utilizado en el diagrama convencional. Los rasgos representados están más repartidos y, por tanto, son más fáciles de interpretar (Wadhwa et al., 2018).

**Observación 1.** *Existen varios paquetes de software que computan homología persistente, por ejemplo Javaplex <http://appliedtopology.github.io/javaplex/> y Ripser <https://pypi.org/project/ripser/>. En este trabajo utilizamos Ripser que implementa un algoritmo sobre una filtración de complejos Vietoris-Rips (Bauer, 2021; Tralie et al., 2018).*

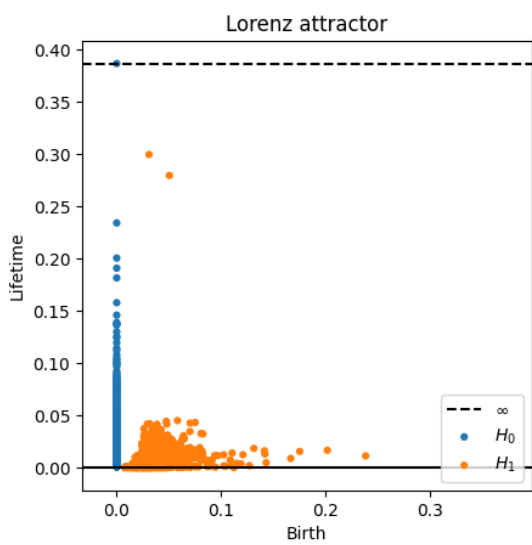
En el siguiente capítulo aplicaremos la teoría de las homología persistentes a un conjunto de series temporales asociadas a variables escalares de cuatro atractores climáticos. Con estos datos construiremos los diagramas de persistencia de los atractores con el objetivo de diferenciarlos topológicamente y así captar las características más significativas.



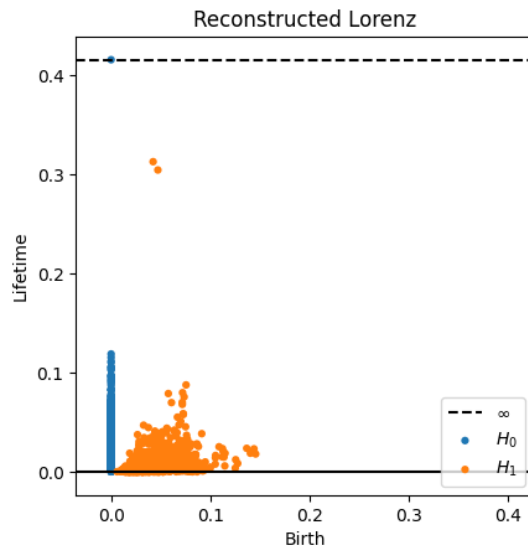
(a) Nube del atractor original



(b) Nube del atractor reconstruido

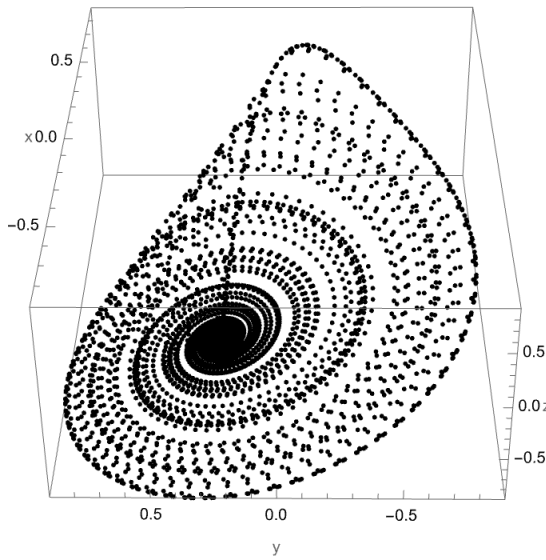


(a) Diagrama del atractor original

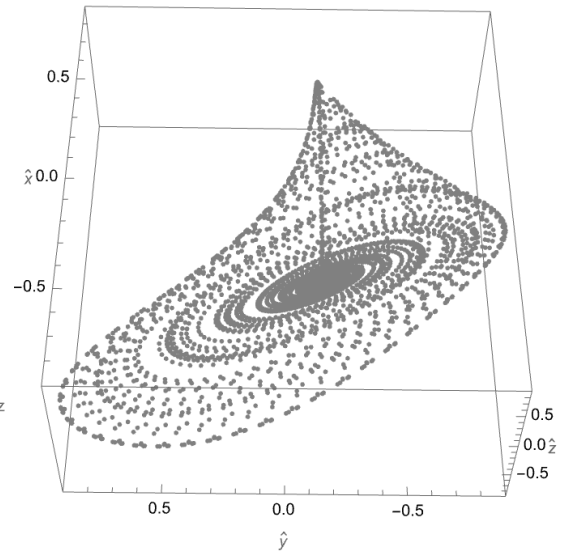


(b) Diagrama del atractor reconstruido

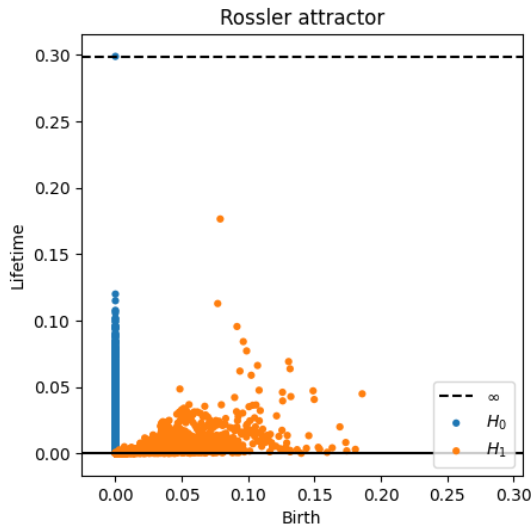
**Figura 2.10:** Nubes y diagramas de persistencia plana del atractor de Lorenz original y reconstruido con M-SSA.



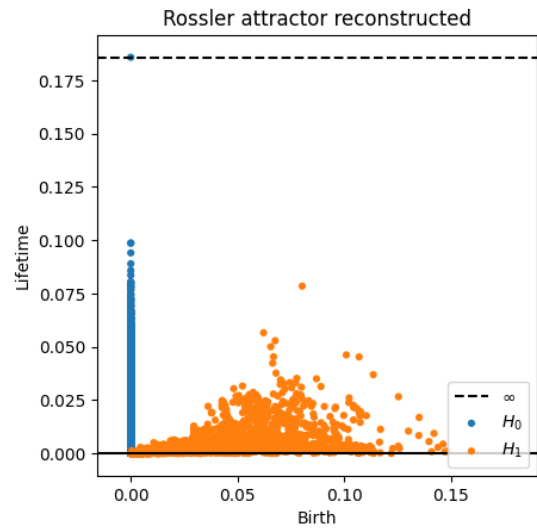
(a) Nube del atractor original



(b) Nube del atractor reconstruido



(a) Diagrama del atractor original



(b) Diagrama del atractor reconstruido

**Figura 2.11:** Nubes y diagramas de persistencia plana del atractor de Rössler original y reconstruido con M-SSA.



## Capítulo 3

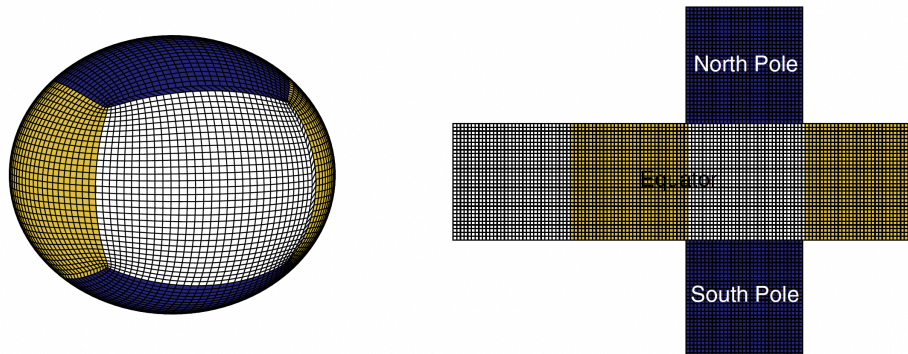
# Análisis topológico de atractores climáticos

El concepto de clima es una de esas abstracciones que parecen evidentes al profano, pero que no son tan sencillas de definir científicamente (Ghil, 1976). La idea intuitiva del clima tiene dos aspectos: (i) las características más importantes de los fenómenos atmosféricos; y (ii) el comportamiento medio de estos fenómenos en un periodo de tiempo convenientemente largo y en zonas convenientemente extensas. Las dificultades comienzan cuando se intenta dar un significado preciso a las palabras claves “más importantes”, “convenientemente largo” y “convenientemente grande”.

En el siglo I a.C., el escritor latino Lucrecio, acuña la frase *quod ali cibus est aliis fuat acre venenum*, es decir, lo que es alimento para un hombre puede ser veneno amargo para otros. En ciencias del clima, se suele decir: “Someone’s noise is another one’s signal”, es decir, el ruido es para uno lo que es señal para otro. Para un climatólogo, el clima es la señal y el tiempo (‘weather’ en inglés) es el ruido, que es a su vez la señal para el meteorólogo. Con el advenimiento de nuevas tecnologías, no solamente instrumentales, como la teledetección satelital, sino también computacionales, como las simulaciones numéricas de alto rendimiento, las observaciones (experimentales o numéricas) permiten trabajar con una resolución espacial y temporal cada vez mayor, dando acceso a procesos que, en muchos casos, son modelados como un ruido subyacente. Ahora bien, ¿qué nivel de resolución espacio-temporal (en el espacio de fases) es “conveniente” utilizar para que un análisis topológico de series temporales climáticas sea exitoso y nos permita distinguir dinámicas cualitativamente diferentes a través de diagramas de persistencia?

Los atractores climáticos que serán sometidos a consideración en este trabajo, son producto de simulaciones numéricas con un modelo de circulación global (GCM, Global Circulation Model, en inglés) del MIT en una grilla definida sobre una esfera cúbica, donde cada cara del cubo tiene  $32 \times 32$  celdas, tal como se muestra en la Figura 3.1. La malla oceánica tiene 15 niveles verticales de distinto espesor, desde 50 *m* cerca de la superficie hasta 690 *m* en el abismo, para una profundidad total de 3 *km*. De este modo, es posible realizar simulaciones de miles de años en un tiempo de CPU razonable (100 años en un día en una estación de trabajo normal).

La descripción que se ofrece de estos atractores en (Brunetti et al., 2019), se basa en técnicas para comprobar la complejidad de las series temporales (Tang et al., 2016) y en el estudio de los modos espaciales de variabilidad de la dinámica climática en cada atractor. El primer estudio utiliza conceptos tales como la dimensión instantánea (Faranda et al., 2017) que cuantifica el número de veces que una trayectoria en el espacio de fases vuelve



**Figura 3.1:** Esfera cúbica utilizada en el modelo MITgcm y su representación crossmap (Brunetti et al., 2019).

sobre una esfera centrada en un punto de un atractor. El segundo, propone un análisis estadístico de componentes principales (Jolliffe and Cadima, 2016). En ambos casos, se trabaja con series temporales promediadas anualmente, en cada punto de la malla horizontal en ventanas de análisis de 1500 años.

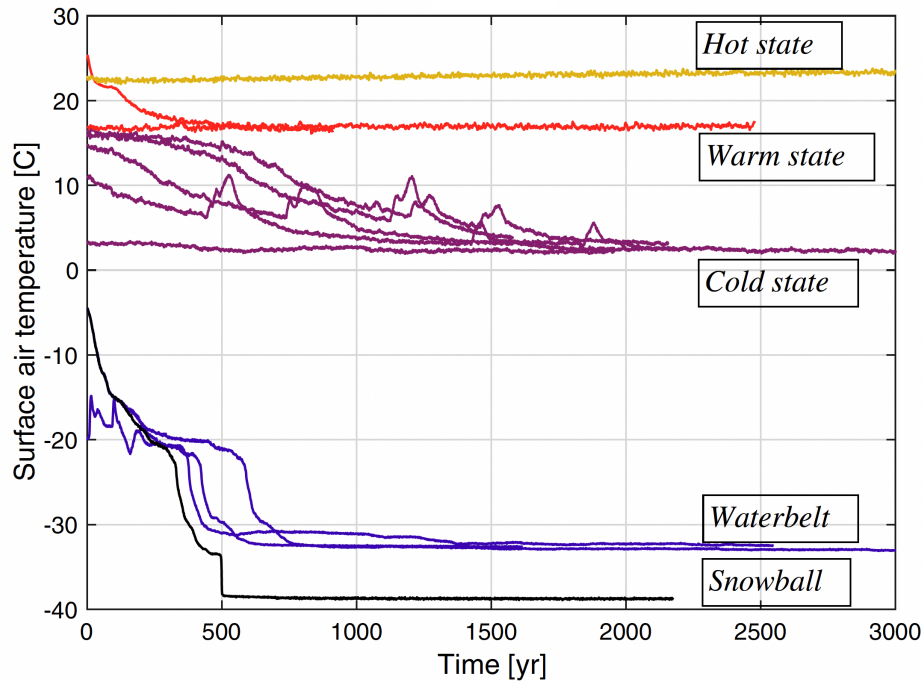
Las series temporales de los promedios anuales resultan, al ser embebidas, indistinguibles entre sí, dando lugar a nubes de puntos distribuidas en forma de esferas o toros. Una esfera sólida en el espacio de fases puede ser vista como la versión estadística de un punto fijo en el espacio de fases: descartado el transitorio, el sistema se estabiliza en el entorno de un punto determinado. Por otra parte, cuando la dinámica que predomina en la variabilidad global de un atractor climático es la del ciclo estacional, las series embebidas se inscriben en un toroide, tal como fue reportado para series temporales diarias obtenidas con diferentes modelos y con reanálisis en el Océano Pacífico entre  $20^{\circ}N$  y  $20^{\circ}S$  desde Enero de 1979 hasta diciembre de 2019 (Falasca and Bracco, 2022).

Ahora bien, es sabido que las propiedades dinámicas de un atractor climático dependen de sus propiedades locales e instantáneas, más que de sus propiedades medias (Vautard and Ghil, 1989; Lucarini et al., 2016). Dado que aspiramos a distinguir a los atractores por sus dinámicas particulares a través de propiedades topológicas diferenciadas en el espacio de fases, deberemos utilizar series temporales con una resolución temporal inferior a la utilizada en el trabajo de referencia.

El análisis topológico con homologías persistentes de los datos climáticos que presentaremos en este capítulo concierne las propiedades topológicas de embebidos que describen la evolución de cuatro atractores en ventanas temporales deslizantes de 1000 días (en un rango de 5000 días) y con resolución temporal diaria. Las series temporales fueron calculadas por Maura Brunetti específicamente para esta ocasión. La exploración concierne un conjunto de campos escalares con variables oceánicas y atmosféricas para una grilla rectangular longitud  $\times$  latitud de  $180 \times 90$ . La técnica de análisis singular de espectros multi-canal descrita en la sección (1.6) será utilizada para preprocesar las series temporales asociadas a un mismo atractor climático.

Para analizar los diagramas de persistencia, se definirá un conjunto de ocho marcadores o índices topológicos que permiten condensar las propiedades de los diagramas y compararlos entre sí.





**Figura 3.2:** Evolución de la temperatura del aire en superficie para los distintos atractores estudiados en Brunetti et al. (2019): transitorio y estado estacionario.

### 3.1. Simulaciones con modelos climáticos

En (Brunetti et al., 2019) se estudian los atractores climáticos que resultan de utilizar condiciones iniciales (CI) construidas del siguiente modo. El océano se inicializa a partir de un estado de reposo, con una temperatura oceánica media de  $8,9^{\circ}\text{C}$ , y se realizan simulaciones con distintos valores del umbral de humedad relativa para las nubes bajas, de modo que se obtienen condiciones extremadamente frías o cálidas. El conjunto de CI resultantes abarcan medias globales iniciales de la temperatura del aire en superficie (*sat*, surface air temperature en inglés) que oscilan entre  $-40$  y  $35^{\circ}\text{C}$ . La Figura 3.2 muestra la evolución temporal de *sat* y cómo diferentes condiciones iniciales convergen hacia cinco atractores diferenciados, que van desde una bola de nieve hasta un estado caliente sin hielo marino.

Cada atractor corresponde a un clima cualitativamente diferente, que es el resultado de la competencia entre diversos mecanismos no lineales e interacciones entre el océano, la atmósfera y el hielo marino. Se los denomina: *Hotstate*, un estado cálido en el que el hielo marino desaparece por completo, *Warmstate*, un estado templado similar a nuestro planeta, con temperaturas intermedias, *Coldstate*, un un estado frío donde el hielo marino se extiende hasta latitudes medias, *Waterbelt*, un cinturón de agua en el Ecuador. Existe un quinto estado posible, denominado *Snowball* en el que todo el planeta está cubierto de hielo. En este último atractor, no hay modos que describan más del 4% de la varianza total, mostrando la falta de una dinámica espacial marcada en este régimen (Brunetti et al., 2019), razón por la cual no lo incluiremos en el análisis topológico.

La tabla de la Figura 3.3 presenta los valores medios globales de la temperatura del aire en superficie, la temperatura oceánica, la extensión del hielo marino, la latitud del límite del hielo marino, el balance en el tope de atmósfera y en la superficie del océano, y

Attractor	Waterbelt	Cold state	Warm state	Hot state
SAT (°C)	$-33.00 \pm 0.03$	$2.2 \pm 0.1$	$17.0 \pm 0.2$	$23.4 \pm 0.1$
Ocean temperature (°C)	$-1.6449 \pm 0.0008$	$3.225 \pm 0.002$	$9.877 \pm 0.004$	$17.418 \pm 0.006$
Sea ice extent ( $10^6 \text{ km}^2$ )	$429.51 \pm 0.06$	$160.4 \pm 0.9$	$67 \pm 1$	0
Latitude of sea ice boundary	9	43	60	90
TOA budget ( $\text{W/m}^2$ )	$1.6 \pm 0.1$	$3.0 \pm 0.2$	$2.6 \pm 0.2$	$2.3 \pm 0.2$
Ocean surface budget $Q_{net}$ ( $\text{W/m}^2$ )	$0.01 \pm 0.14$	$-0.009 \pm 0.217$	$0.03 \pm 0.29$	$0.05 \pm 0.27$

**Figura 3.3:** Valores medios globales de los distintos atractores promediados a lo largo de los últimos 100 años de simulaciones realizadas en (Brunetti et al., 2019).

el desequilibrio de la superficie oceánica del flujo de radiación para los cuatro atractores climáticos de los que nos ocuparemos en lo que queda del capítulo.

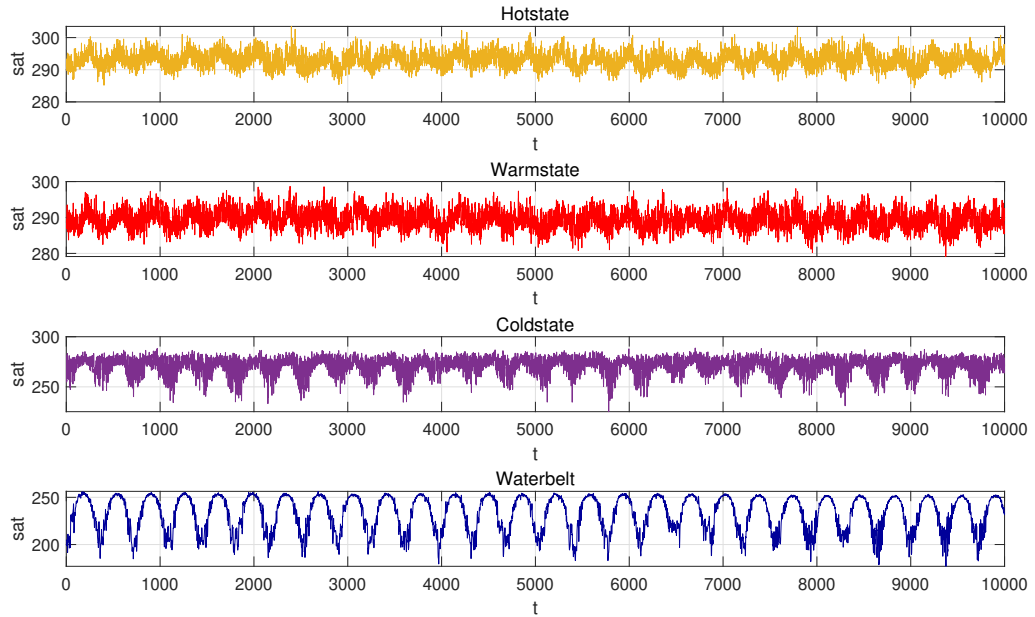
## 3.2. Series temporales atmosféricas y oceánicas

Numerosos estudios climáticos se basan en la perspectiva que considera que cada registro local es posible de ser visto como un oscilador caótico que representa la dinámica regional. Es más, el atractor de Rössler presentado en el primer capítulo, también llamado oscilador de Rössler o ‘Rösselador’ – el sistema caótico más simple que pueda concebirse – ha sido utilizado como prototipo para estudiar al sistema climático como una cadena de osciladores acoplados por difusión (Groth et al., 2017).

Métodos como el M-SSA permiten examinar las propiedades espectrales cruzadas de los registros climáticos de diversas variables en diferentes regiones, a través de un puñado de modos oscilatorios significativos en una multiplicidad de canales (cada variable en un punto de grilla constituye un canal). Utilizando los componentes principales del M-SSA para reconstruir un conjunto de series temporales, se obtiene una serie temporal sustituta, menos ruidosa que las originales pero con relevancia estadística. El resultado obtenido permite obtener un ‘esqueleto’ robusto de la estructura del sistema dinámico bajo estudio. La dimensión del espacio de fases para este esqueleto es, desde ya, inferior a la original, aportando una visión simplificada del sistema, que busca retener, no obstante, los mecanismos no lineales preponderantes.

La mezcla de variables asociada al análisis multicanal puede reducir el poder explicativo del M-SSA. Por esta razón, es una práctica frecuente hacer análisis separados por tipo de variable (Groth et al., 2017), para obtener series temporales sustitutas consistentes, representativas de cada subsistema. Nosotros trabajaremos con cuatro puntos de grilla elegidos como representativos del planeta en su conjunto y con dos tipos de variables escalares:

- variables atmosféricas: temperatura del aire en superficie (sat), humedad específica del aire en la superficie (qas), y precipitación (prr).
- variable oceánica: temperatura de la superficie del mar (sst).



**Figura 3.4:** Serie temporal de la variable atmosférica sat (temperatura de la superficie del aire), sin ningún pre-procesamiento, para los cuatro atractores climáticos a estudiar.

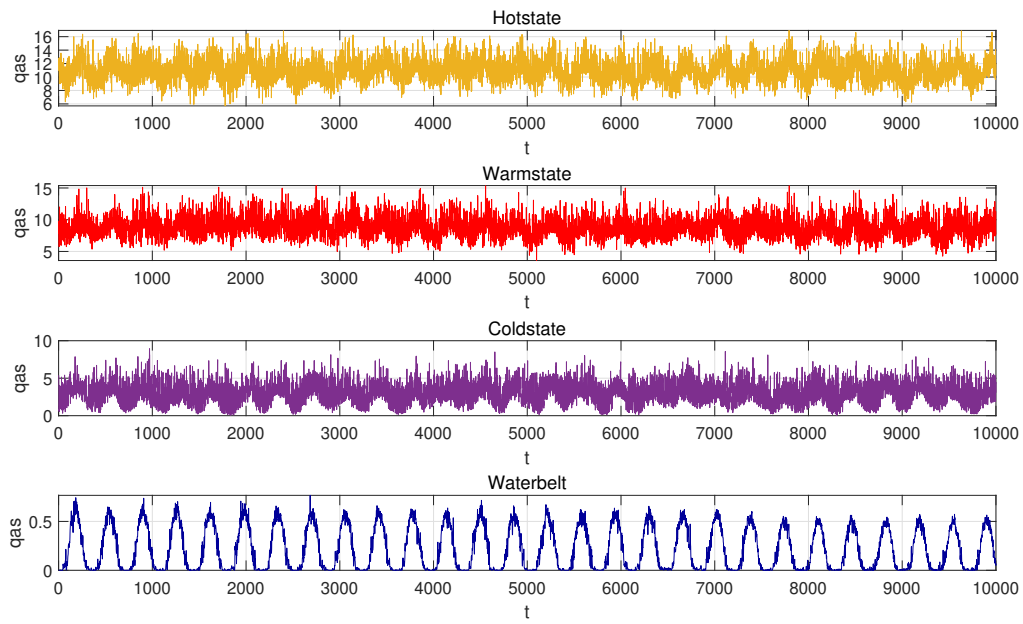
El método multicanal de análisis espectral nos permitirá en consecuencia obtener una sola serie temporal simplificada, representativa del subsistema atmósfera para cada atractor climático, y una sola serie simplificada del subsistema océano. La técnica de embebidos permite a su vez reconstruir con ellas una nube de puntos, utilizando retardos temporales  $\tau_{a,v}$  estimados con AMI, donde  $a$  denota el atractor y  $v$  el tipo de variable, y una dimensión mínima  $n_{min}$  del espacio de fases reconstruido, que en principio también puede variar con el atractor y con el tipo de variable utilizada.

### 3.2.1. Pre-procesamiento

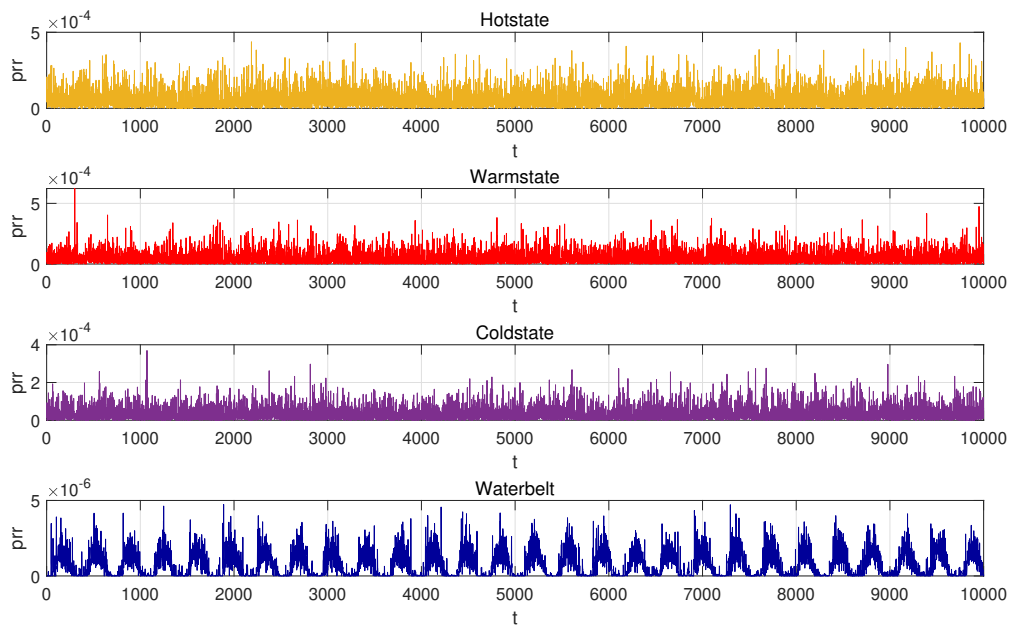
En las Figuras 3.4, 3.5, 3.6, y 3.7 se observan las series temporales originales para las variables atmosféricas (sat, gas, prr y sst) en el punto ( $61^{\circ}N, 59^{\circ}E$ ) para cada atractor. En todas estas figuras se observa una periodicidad característica de los ciclos estacionales. Los patrones estacionales son importantes, pero ya conocidos. Con el objetivo de encontrar patrones no estacionales, eliminamos el ciclo estacional de los datos. En las Figuras 3.8 y 3.9, se presentan, la variable atmosférica sat (temperatura del aire en superficie) y la variable oceánica sst (temperatura de la superficie del mar) una vez filtrado el ciclo estacional. Trabajaremos a continuación con todas las variables preprocesadas de esta manera.

Con el objeto de lograr una reducción dimensional del espacio de fases, se aplica la técnica M-SSA, descrita en la sección 1.6. Para la implementación del M-SSA, se utiliza el paquete *Melissa* desarrollado por Mario Sandri (Sandri, 2003) para Wolfram Mathematica © y actualizado en <https://git.cima.fcen.uba.ar/sciamarella/melissa-update>. El código se ejecuta en el entorno gratuito de Jupiter Notebook a través del kernel <https://github.com/WolframResearch/WolframLanguageForJupyter> (Wolfram Engine).

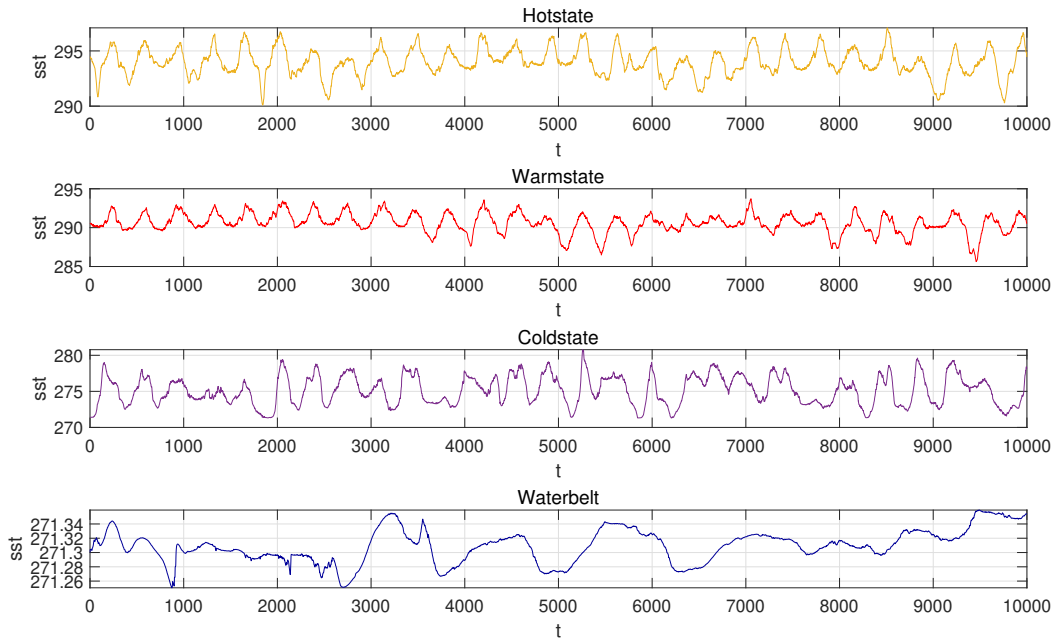
Se consideraron como canales a las diferentes variables climatológicas en cuatro puntos



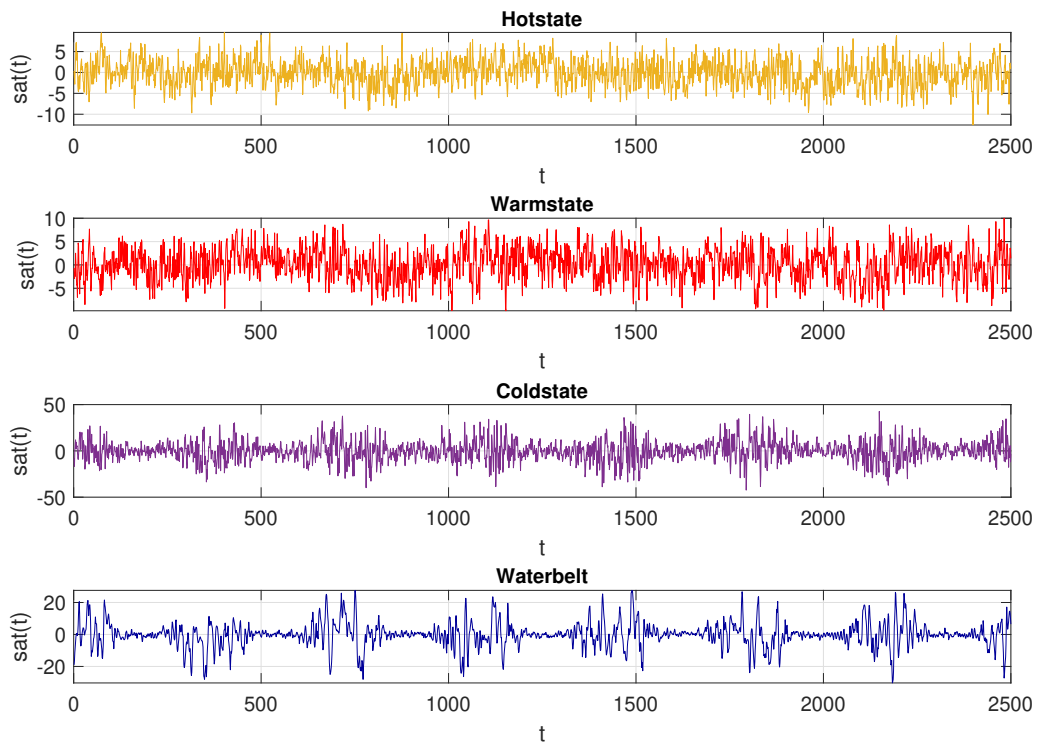
**Figura 3.5:** Serie temporal de la variable atmosférica gas, sin ningún pre-procesamiento, para los cuatro atractores climáticos a estudiar.



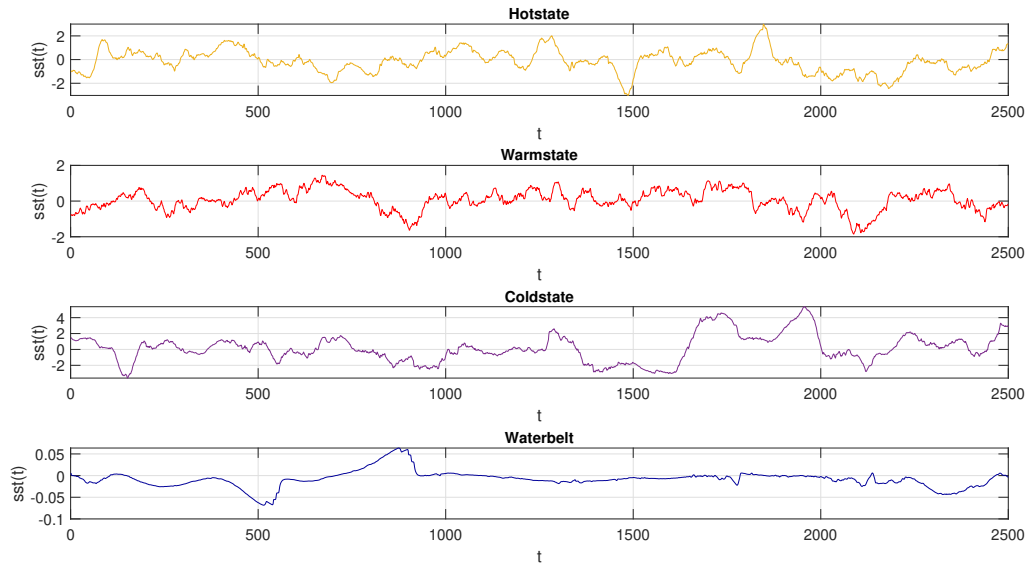
**Figura 3.6:** Serie temporal de la variable atmosférica prr, sin ningún pre-procesamiento, para los cuatro atractores climáticos a estudiar.



**Figura 3.7:** Serie temporal de la variable oceánica sst, sin ningún pre-procesamiento, para los cuatro atractores climáticos a estudiar.



**Figura 3.8:** Serie temporal de la variable sat (temperatura del aire en superficie) una vez filtrado el ciclo estacional.



**Figura 3.9:** Serie temporal de la variable sst (temperatura de la superficie del mar) una vez filtrado el ciclo estacional.

representativos de las zonas del planeta: latitudes medias ( $-21^{\circ}N, 59^{\circ}E$ ) y ( $41^{\circ}N, 59^{\circ}E$ ), Ecuador ( $-1^{\circ}N, 59^{\circ}E$ ) y polos ( $61^{\circ}N, 59^{\circ}E$ ). En el primer análisis M-SSA se usaron como canales a las 12 series temporales correspondientes a las variables atmosféricas, y en el segundo se usaron las 4 series temporales correspondientes a la única variable oceánica, sst.<sup>1</sup>

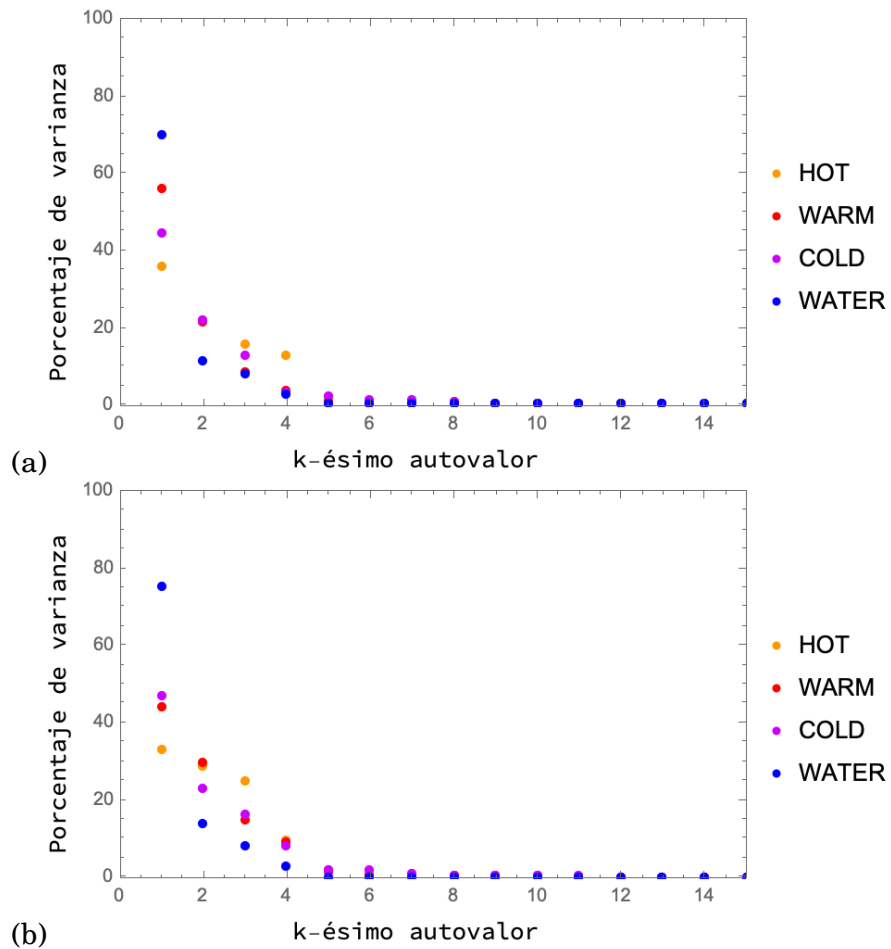
Cumpliendo con los requerimientos para seleccionar el valor de  $M$  detallados en el capítulo 1.6, se fijó una ventana de  $M = 60$  para construir la matriz de trayectoria multicanal. De acuerdo a lo establecido para el análisis separado por tipo de variable, se utilizan 12 canales para las variables atmosféricas y 4 para las variables oceánicas. Los porcentajes de varianza asociados a los autovalores para cada caso se muestran en la Figura 3.10. En ambos gráficos se observa que los primeros 6 autovalores concentran casi toda la varianza de la muestra. Los resultados de estos análisis permiten aproximar las series temporales atmosféricas y oceánicas con una reconstrucción ( $Y_a$  e  $Y_o$ , respectivamente) que usa los 6 primeros RCs del análisis multicanal. En las Figuras 3.11 y 3.12 se observan las series  $Y_a(t)$  e  $Y_o(t)$ , que es la suma de las reconstrucciones usando los 6 primeros autovectores para cada atractor.

### 3.2.2. Nubes de puntos

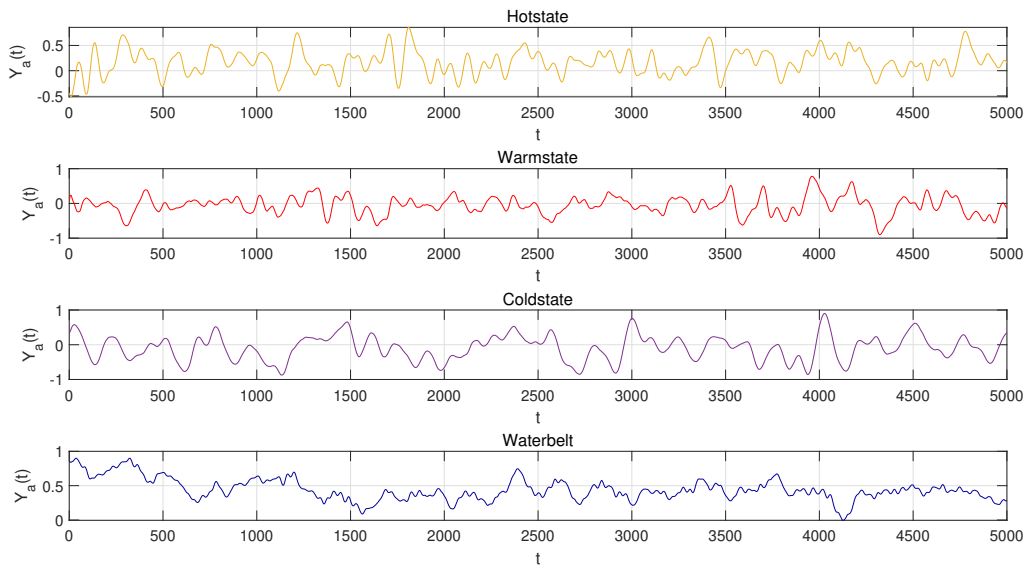
El objetivo final de este análisis es distinguir atractores climáticos a partir de nubes de puntos que constituyen una reconstrucción parcial de su dinámica oceánica y atmosférica. El pre-procesamiento realizado excluye la caracterización de propiedades topológicas vinculadas al ciclo estacional o a la varianza asociada a las componentes que fueron excluidas del análisis al retener los primeros seis autovalores.

Las nubes de puntos que se someten al análisis topológico son embebidos por retardos

<sup>1</sup>Cabe aclarar que el procedimiento habitual, cuando se dispone de una grilla de datos, consiste en proyectar la grilla completa ponderada por el coseno de la latitud, sobre un subconjunto de modos espaciales mediante un análisis convencional de componentes principales. Esta compresión reduce el costo computacional del análisis M-SSA subsiguiente (Dettinger et al., 1995; Moron et al., 1998; Groth et al., 2017).

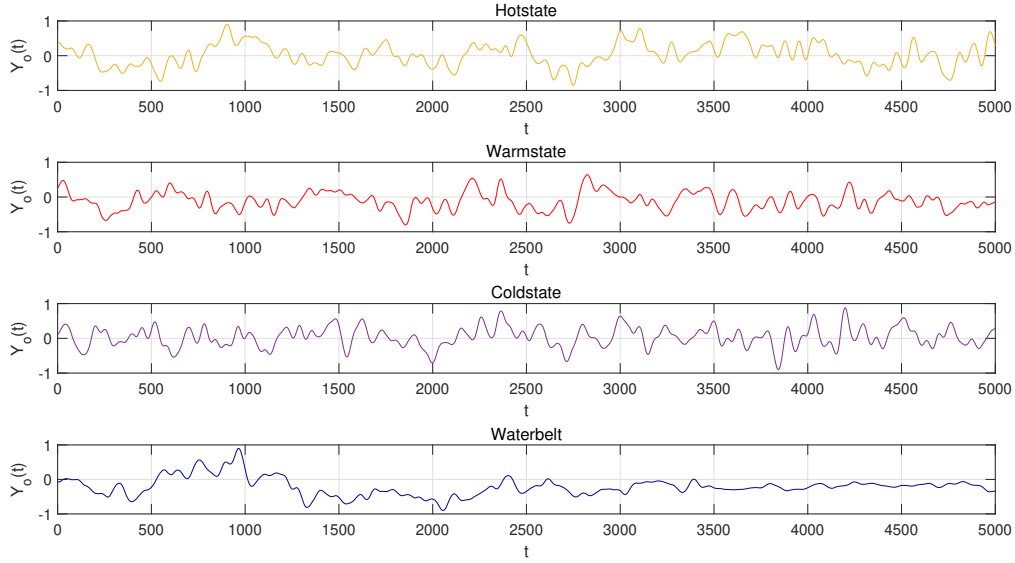


**Figura 3.10:** Porcentajes de varianza asociados a los autovalores de las variables atmosféricas (arriba) y para las variables oceánicas (abajo).



**Figura 3.11:** Series reconstruidas usando los 6 primeros RCs para cada atractor.





**Figura 3.12:** Series reconstruidas  $Y_o$  usando los 6 primeros RCs para cada atractor.

temporales (ver sección 1.5) utilizando los primeros 1000 días de cada una de las series reconstruidas en la sección anterior, en las Figuras 3.11 y 3.12.

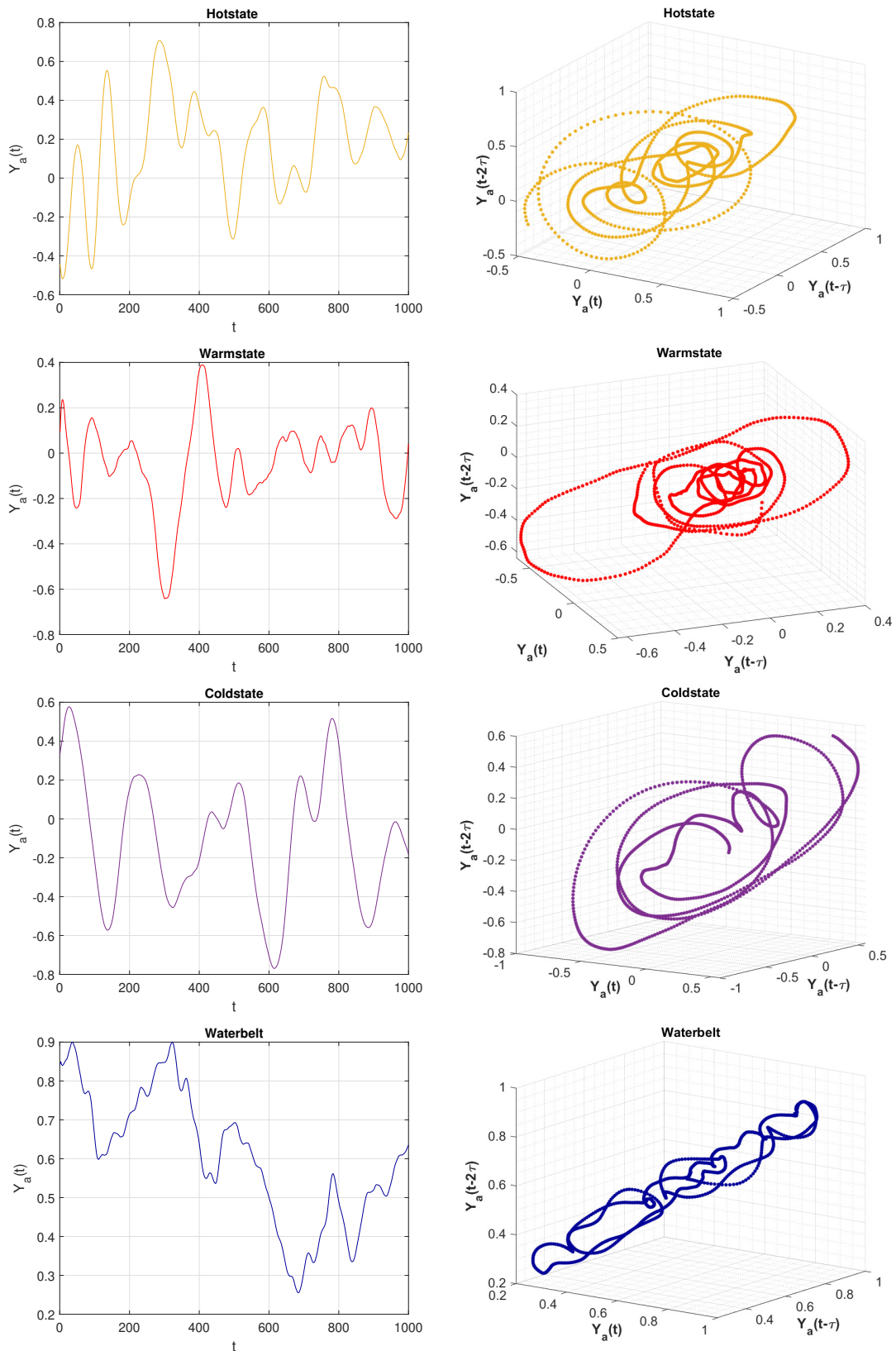
Con el objetivo de reconstruir el espacio de fases que representa parcialmente a cada atractor, se tomó la ventana temporal  $T_w = [0, 1000]$  y se aplicó el algoritmo NoLitsa a cada embebido con retardo temporal  $\tau = 14$ ,  $\tau = 16$ ,  $\tau = 10$  y  $\tau = 20$ , correspondientes a los atractores del hotstate, warmstate, coldstate y waterbelt, respectivamente. La dimensión que surge del resultado de NoLitsa aplicado a los cuatro embebidos es de 3 para todos los atractores. Desde ya, esto no significa que los datos originales tengan esta dimensión. Es un resultado que está sujeto al pre-procesamiento de las series y a su reconstrucción, teniendo también en cuenta las ventanas temporales utilizadas para construir el embebido. El resultado obtenido expresa que tres dimensiones son suficientes (no existen auto-cruces) para caracterizar topológicamente la dinámica contenida en las series temporales reconstruidas con 1000 días.

En las Figuras 3.13 y 3.14 se muestran las series temporales y sus correspondientes embebidos en  $\mathbb{R}^3$  para las variables atmosféricas y oceánicas, respectivamente.

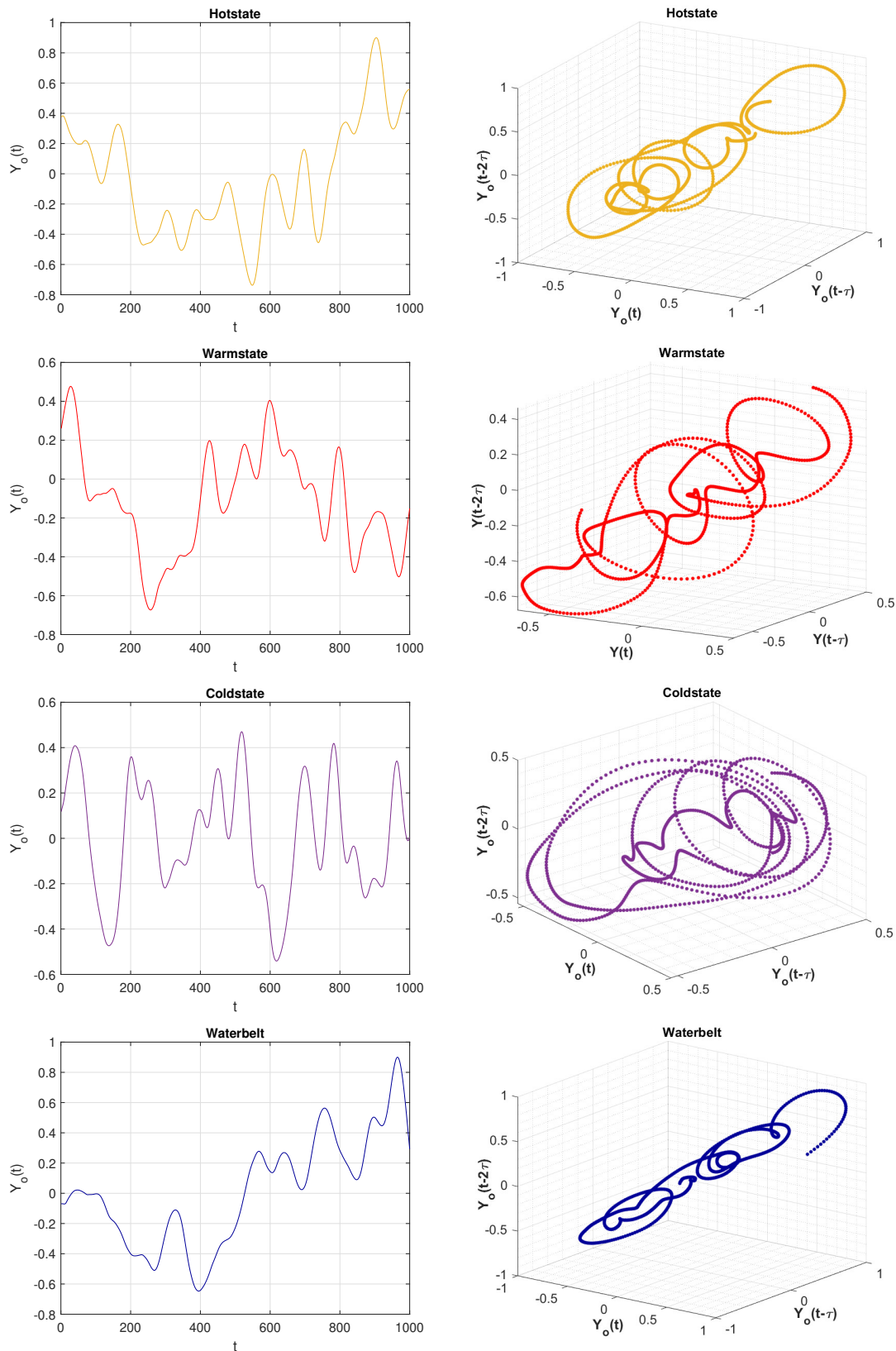
### 3.3. Homologías persistentes y atractores climáticos

El empleo de diagramas de persistencia para establecer marcadores o índices topológicos ha sido previamente utilizado para clasificar proteínas (Cang et al., 2015), apelando a la construcción de vectores con valores de una selección de índices o marcadores que permiten seleccionar rasgos distintivos. En esta sección, presentaremos los diagramas de persistencia para las nubes de puntos obtenidas en la sección anterior. Luego procederemos a la construcción de marcadores relacionados con los intervalos de Betti de los diagramas, y por último, haremos un análisis comparativo de los cuatro atractores en términos de los marcadores topológicos previamente definidos.

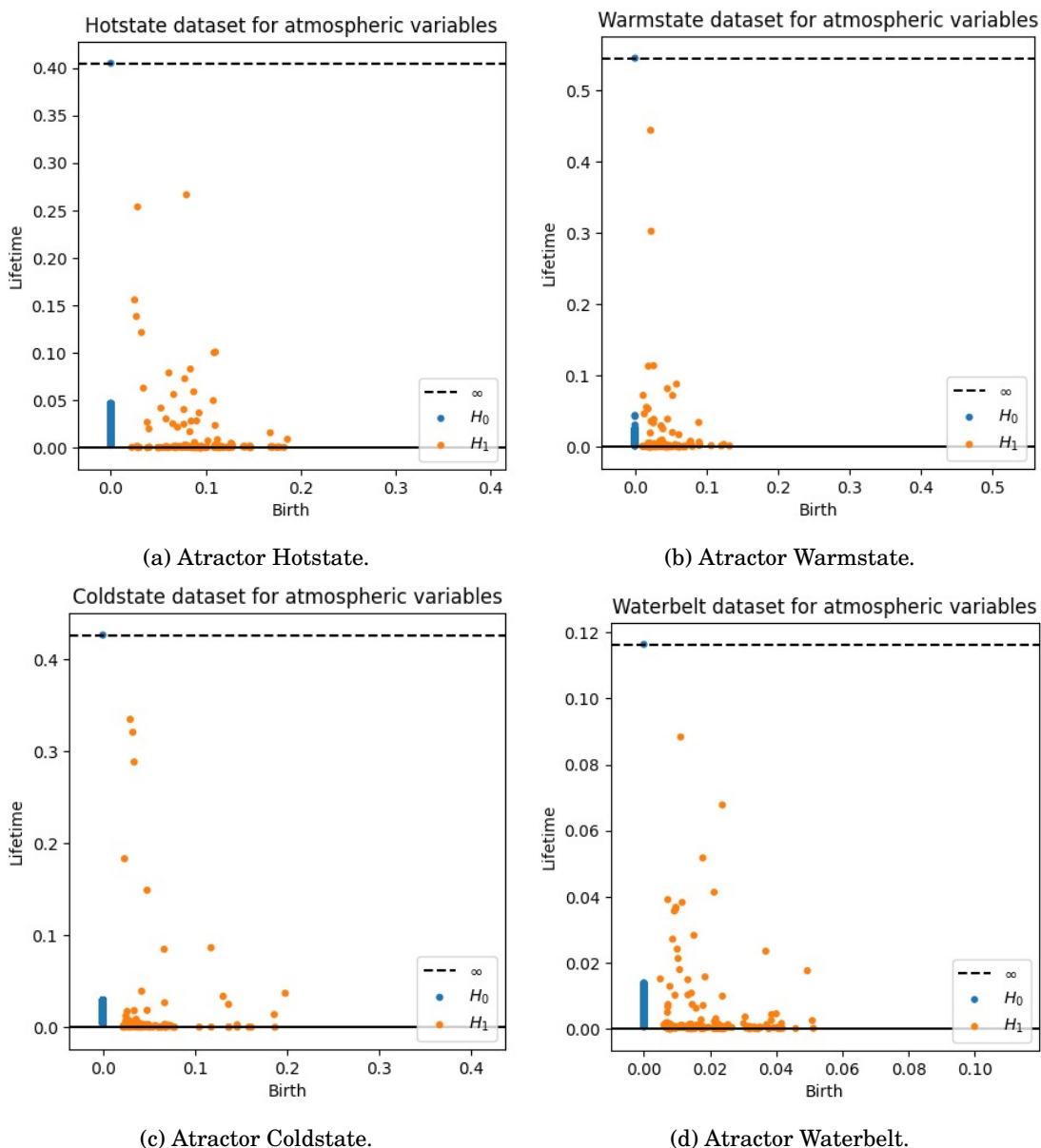




**Figura 3.13:** Variables atmosféricas: A la izquierda se observan las series temporales de cada atractor para las reconstrucciones de la Figura 3.11 en la ventana temporal  $T_W = [0, 1000]$ . A la derecha se muestran los embebidos de retardo temporal:  $\tau = 14$  (Hotstate),  $\tau = 16$  (Warmstate),  $\tau = 10$  (Coldstate) y  $\tau = 20$  (Waterbelt).



**Figura 3.14:** Variable oceánica: A la izquierda se observan las series temporales de cada atractor para las reconstrucciones de la Figura 3.12 en la ventana temporal  $T_W = [0, 1000]$ . A la derecha se muestran los embebidos de retardo temporal:  $\tau = 14$  (Hotstate),  $\tau = 16$  (Warmstate),  $\tau = 10$  (Coldstate) y  $\tau = 20$  (Waterbelt).

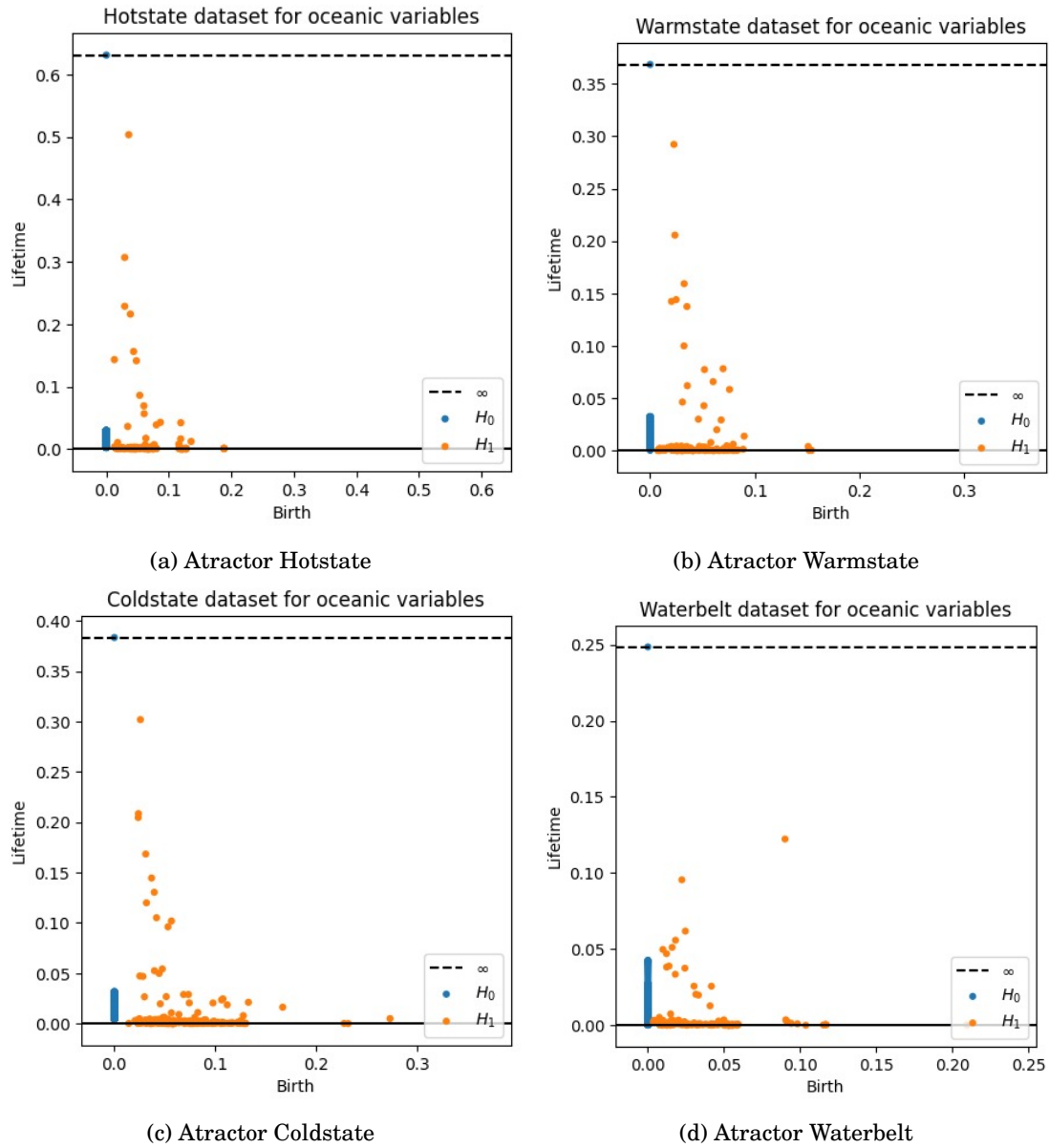


**Figura 3.15:** Diagramas de persistencia de los atractores climáticos con las variables atmosféricas para  $H_0$  y  $H_1$ .

### 3.3.1. Diagramas de persistencia

Tal como se indica en 2.4.2, el cómputo de los diagramas de persistencia correspondientes a las nubes de puntos obtenidas en la sección 3.2.2 se realiza con el paquete Ripser para Python en el que se implementa la filtración Vietoris Rips. En la Figura 3.15 se observan los diagramas de persistencia correspondientes a las variables atmosféricas y en la Figura 3.16 los correspondientes a las variables oceánicas.

El tiempo de cómputo de los diagramas depende del tamaño de la nube de puntos en particular y de la dimensión de los grupos de homología que se pretende calcular. Con este fin, se utilizó una máquina virtual del servicio gratuito de Google cloud llamado Compute engine, la vm elegida fue una c2-standard-4. Se comenzó trabajando con  $H_0$ ,  $H_1$  y  $H_2$ , pero dado que los resultados que permiten discriminar atractores están mayormente asociados a  $H_0$  y  $H_1$ , los diagramas que se presentan usan una dimensión máxima de 1.



**Figura 3.16:** Diagramas de persistencia de los atractores climáticos para las variables oceánicas para  $H_0$  y  $H_1$ .

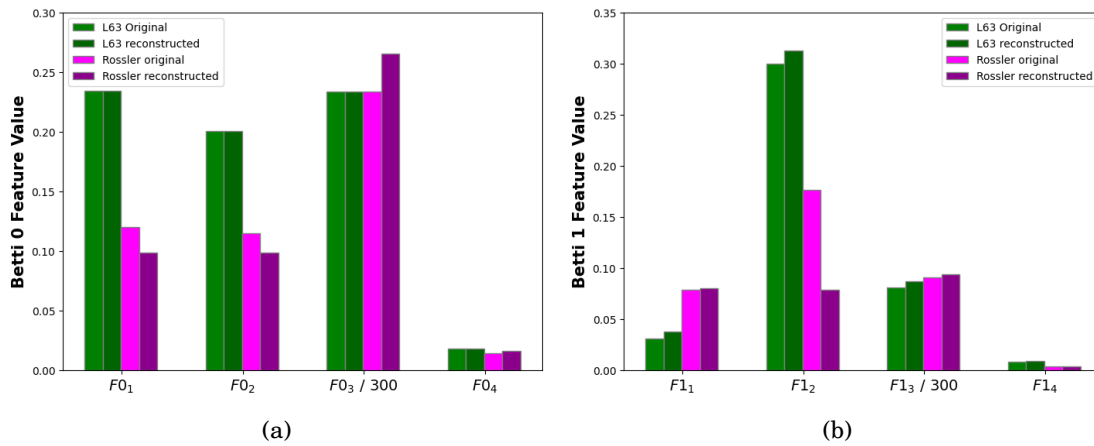
### 3.3.2. Marcadores topológicos

La lista a continuación presenta a los marcadores topológicos (*markers* o también *features*, en inglés) basados en los intervalos de Betti 0 y 1 de los diagramas de persistencia. Las definiciones provienen de (Cang et al., 2015). El *script* de Python que se utiliza para calcularlos está disponible en <https://git.cima.fcen.uba.ar/sciamarella/ph-for-climate-attractors>.

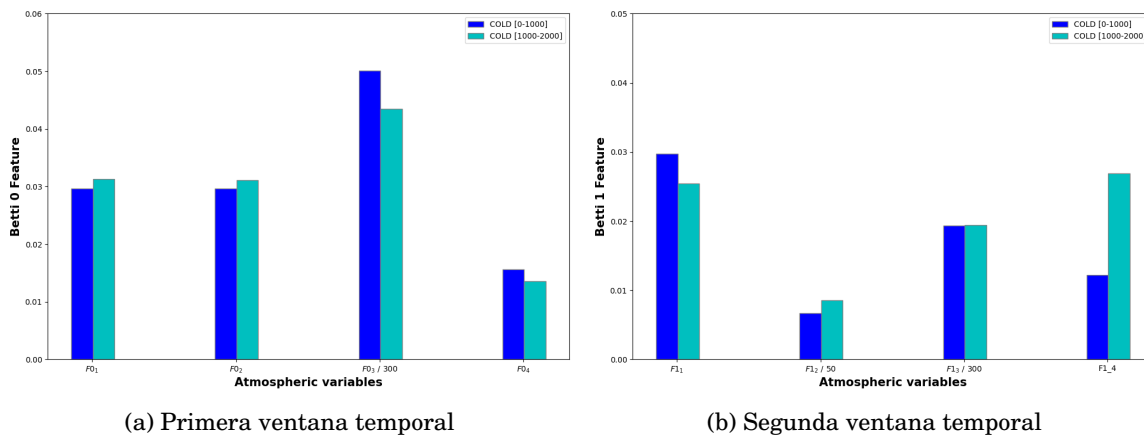
- $F0_1$ : longitud del segundo intervalo más largo de Betti 0 (indica el inicio en la filtración donde los simplices del complejo correspondiente forman una componente conexa).
- $F0_2$ : longitud del tercer intervalo más largo de Betti 0 (indica el inicio en la filtración donde los simplices forman dos componentes conexos). En general, toma valores cercanos a  $F0_1$ .
- $F0_3$ : suma de las longitudes de todos los intervalos de Betti 0 excepto los que superan el valor máximo de filtración (describe el grado de compacidad de los puntos).
- $F0_4$ : longitud media de los intervalos de Betti 0, excepto los que superan el valor máximo de filtración.
- $F1_1$ : el valor de inicio (nacimiento) del intervalo de Betti 1 más largo (muestra el valor de filtración en el que se forma el 1-agujero persistente más grande).
- $F1_2$ : longitud del intervalo de Betti 1 más largo (refleja el tamaño del 1-agujero geoméricamente dominante).
- $F1_3$ : suma de las longitudes de todos los intervalos de Betti 1 excepto los que superan el valor máximo de filtración.
- $F1_4$ : longitud media de los intervalos de Betti 1, excepto los que superan el valor máximo de filtración (indicativo del tamaño medio de los 1-agujeros).

Observemos en la Figura 3.17 qué valores adoptan estos marcadores para los atractores de Lorenz y de Rössler, tanto para las nubes originales como para las reconstruidas con M-SSA. Vemos que en general, las nubes reconstruidas y las originales tienen valores muy cercanos en todos los marcadores. La única excepción es con el marcador  $F1_2$  para el Rössler original y reconstruido. Este marcador refleja el tamaño del 1-agujero geoméricamente dominante. Dado que la reconstrucción conlleva una deformación respecto de la nube original, es posible que este valor cambie, y debe ser interpretado, en consecuencia, teniendo en cuenta que se trata de un rasgo de predominancia geométrica.

Consideremos ahora uno de los atractores, y analicemos los marcadores topológicos en dos ventanas de tiempo consecutivas, con el objeto de evaluar si los marcadores se mantienen en torno a valores estables. El ejemplo que se muestra en la Figura 3.18 es representativo de lo que ocurre también para los otros atractores: no se observan cambios significativos. El marcador que presenta más diferencias entre las dos ventanas temporales es el  $F1_4$  que ofrece una información de carácter predominantemente geométrico. Se puede concluir, pues, que los marcadores topológicos se mantienen muy cercanos entre sí para distintas ventanas de 1000 días en el rango de 5000 días considerados en este trabajo. Un estudio sistemático de estas leves variaciones puede utilizarse, de ser necesario, para calcular una barra de error para los valores de los marcadores.



**Figura 3.17:** Marcadores topológicos correspondientes al atractor de Lorenz (verde) y al de Rossler (magenta) para las nubes originales y para las reconstruidas.



**Figura 3.18:** Marcadores topológicos atmosféricos correspondientes al atractor frío (*Coldstate*) en la ventana de 0 a 1000 y en la ventana de 1000 a 2000.

Intentaremos ahora describir los distintos atractores a partir de los marcadores establecidos. Los rasgos comparados para los distintos atractores y grupos de variables se muestran en la Figura 3.19. Los rasgos más marcados corresponden al estado caliente y al cinturón de agua.

- **Hotstate:** Se distingue por tener el valor relativo más alto de todos los intervalos de Betti 0 y de  $F_{1,3}$  para las variables atmosféricas.
- **Waterbelt:** Se distingue de los otros atractores por tener el valor relativo más alto de  $F_{0,2}$  y de  $F_{1,1}$  para las variables oceánicas. También presenta los valores relativos más bajos para los intervalos Betti 1 en las variables atmosféricas. Y los valores relativos más bajos de  $F_{0,3,4}$  y  $F_{1,2,3,4}$  para la variable oceánica.

Los otros dos atractores pueden diferenciarse también por características propias, pero con rasgos menos pronunciados.

- **Warmstate:** Se distingue por tener valores relativos más altos de  $F_{1,4}$  que los otros atractores en las variables atmosféricas.
- **Coldstate:** Se distingue por tener el valor relativo más bajo de todos los marcadores de Betti 0 en las variables atmosféricas y el valor relativo más alto de  $F_{0,3,4}$  y de  $F_{1,3}$  para la variable oceánica.

En el cuadro que se presenta a continuación se resumen estas características.

		<i>Hot</i>		<i>Warm</i>		<i>Cold</i>		<i>Water</i>	
		Atm	Sea	Atm	Sea	Atm	Sea	Atm	Sea
<i>Betti 0</i>	$F_{0,1}$	↑				↓			↑
	$F_{0,2}$	↑				↓			↑
	$F_{0,3}$	↑				↓	↑		↓
	$F_{0,4}$	↑				↓	↑		↓
<i>Betti 1</i>	$F_{1,1}$	↑						↓	↑
	$F_{1,2}$			↑				↓	↓
	$F_{1,3}$	↑					↑	↓	↓
	$F_{1,4}$			↑				↓	↓

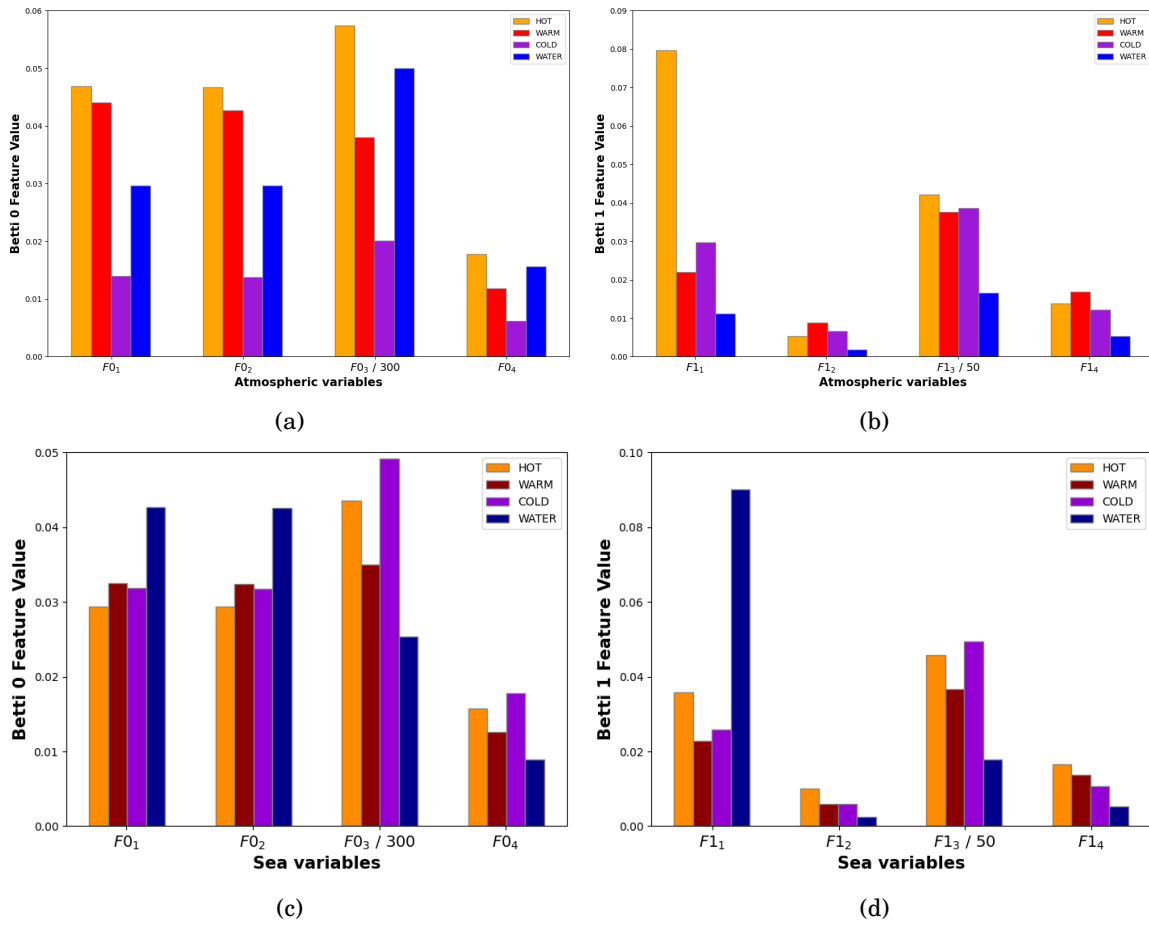
**Cuadro 3.1:** Marcadores topológicos.

En la Figura 3.20 se reúnen los marcadores para cada atractor en una serie de cuatro *heatmaps*.

### 3.4. Discusión

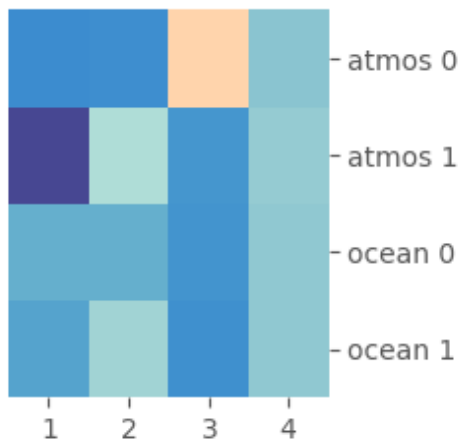
El análisis realizado muestra que cada atractor puede caracterizarse por diferentes valores relativos de los intervalos de Betti en un diagrama de persistencia homológica. Esta posibilidad de discriminar propiedades, está favorecida por el hecho de haber realizado un análisis separado de las variables atmosféricas y de las variables oceánicas, utilizando una técnica avanzada de análisis espectral multicanal que permiten trabajar con series simplificadas mezclando datos obtenidos en varios puntos representativos de la grilla planetaria.

La posibilidad de establecer estas diferencias para un mismo forzante (representado aquí por la misma radiación solar entrante y el mismo contenido atmosférico de dióxido

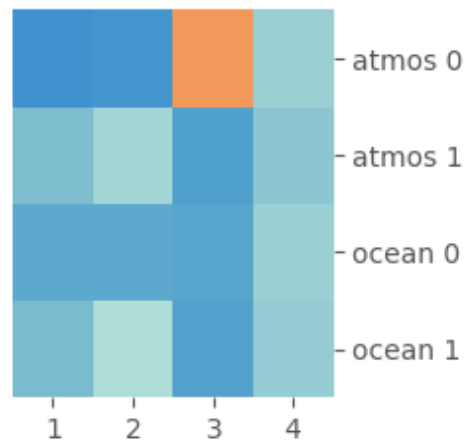


**Figura 3.19:** Marcadores topológicos correspondientes a las variables atmosféricas (a) y (b), y oceánicas (c) y (d) relacionados al intervalo Betti 0 y Betti 1 respectivamente.

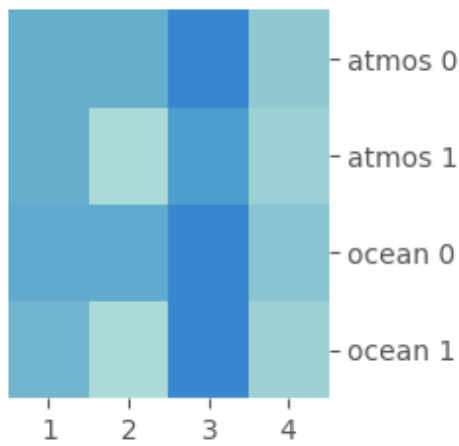




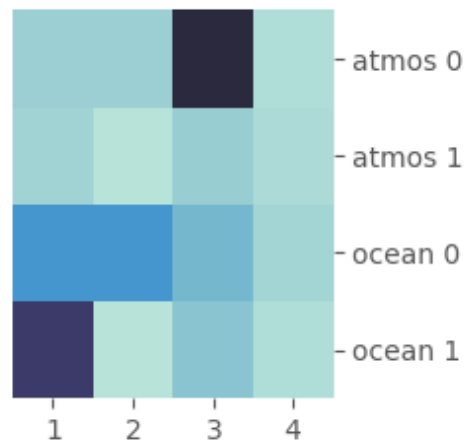
(a) Hotstate



(b) Warmstate



(c) Coldstate



(d) Waterbelt

**Figura 3.20:** Presentación de los cuatro atractores utilizando los cuatro marcadores para los intervalos de Betti 0 y Betti 1 de las variables atmosféricas y oceánicas.

de carbono) puede explicarse en términos de los procesos no lineales que tienen lugar entre el océano, la atmósfera y el hielo marino. Estos procesos pueden equilibrarse de forma diferente en los atractores coexistentes, lo que implica que cada uno de ellos puede identificarse por un conjunto de rasgos diferentes.

Estas diferencias fueron establecidas utilizando promedios de variables climáticas en (Ragon et al., 2022). En particular, la temperatura global, la precipitación total y el pico de intensidad en el transporte de masa de agua desde el ecuador a las regiones polares resultan ser máximos en el estado cálido. El gradiente meridional de temperatura, que afecta al transporte de energía tanto en el océano como en la atmósfera, presenta los picos más destacados en el estado frío. El estado de cinturón de agua se caracteriza por fuertes gradientes meridionales de temperatura y velocidad del viento localizados en los bordes del hielo, que dan lugar a valores elevados de energía potencial disponible y energía cinética en la atmósfera. Estos fuertes gradientes meridionales también pueden explicar la mayor dimensión instantánea del atractor del cinturón de agua con respecto a los fríos, templados y cálidos (Brunetti et al., 2019), en concordancia con (Lucarini and Bódai, 2017; Faranda et al., 2019), donde se demuestra que los gradientes de temperatura fuertes y localizados están asociados a una baja predictibilidad.

En esta tesis mostramos una estrategia alternativa para distinguir estos atractores, ya no en términos de los valores promediados de las variables climáticas sino en términos del modo en que se estructura su evolución temporal para escalas de tiempo de mil días, una vez filtrado el ciclo estacional. Las diferencias en la estructura topológica asociada a las dinámicas propias de las variables atmosféricas u oceánicas de cada atractor, determinada por la forma en que se combinan los procesos no lineales en cada uno de ellos, se vuelven observables pre-procesando las series temporales de las diversas variables, con el fin de obtener una reconstrucción parcial del espacio de fases a tiempo finito con una dimensión baja y poder caracterizar los rasgos salientes en términos de homologías persistentes. Los embebidos por retardos temporales dan lugar a nubes de puntos con características diferenciables.

Utilizando esta misma estrategia con atractores paradigmáticos conocidos, como el Rössler o el Lorenz, se observa que la mariposa de Lorenz tiene un valor más alto del marcador  $F_{0,2}$ . Esto puede atribuirse al hecho de que la formación de un único componente conexo requiere un intervalo de Betti 0 más largo (ligado a la existencia de dos focos repulsores) que en el caso de una única espiral (un único foco repulsor) presente en el Rössler. Este rasgo se observa también para las variables atmosféricas del Hotstate y para las variables oceánicas del Waterbelt, respecto de los otros atractores. Sus espacios de fases reconstruidos requieren estructuras que podríamos llamar multifocales (coexistencia de varios focos) en los subsistemas atmósfera/océano respectivos. En las variables atmosféricas, el Coldstate se estructura en torno a un número relativo más escaso de focos en las variables atmosféricas, en concordancia con sus marcadores topológicos. El Warmstate resulta una especie de categoría intermedia, en la que cobran importancia marcadores como  $F_{1,2}$  y  $F_{1,4}$ , que establecen diferencias entre los tamaños de los 1-agujeros dominantes. Estos marcadores, no obstante, traducen una información geométrica junto con la información topológica. Podríamos denominarlos marcadores mixtos, por su carácter geométrico-topológico. La diferenciación en términos de marcadores mixtos es menos robusta, puesto que si bien pueden ser útiles a los fines de clasificar nubes calculadas de modo tal que las diferencias geométricas no resulten relevantes, pueden resultar confusos a la hora de establecer equivalencias topológicas genuinas.

## Capítulo 4

# Conclusiones y perspectivas

La detección de propiedades características de un sistema dinámico complejo como el Planeta Tierra requiere, en primera instancia, identificar regiones en el espacio de fases hacia las cuales evolucionan asintóticamente las trayectorias solución de un conjunto de ecuaciones que simulan la evolución de las distintas variables climáticas; dichas regiones son conocidas como ‘atractores’. La multiestabilidad que caracteriza el sistema climático terrestre complica el escenario, pues es sabido que, para un mismo forzante, es posible que co-existan atractores de características disímiles.

El objetivo de esta tesis ha sido explorar la posibilidad de distinguir estos atractores coexistentes en un acuplaneta acoplado. El estudio se basa en simulaciones numéricas realizadas con un modelo de circulación general global, teniendo en cuenta la atmósfera y el océano en ausencia de tierra. La complejidad se reduce respecto del clima real terrestre, en un modo en el que no obstante, las interacciones no lineales entre los componentes del clima de un planeta convocan nuestra atención.

Antes de este trabajo, la caracterización de cada uno de los atractores que resultan la dinámica de un acuplaneta había sido realizada a través de una descripción de los valores medios de cantidades globales que dan una idea de las propiedades dominantes en juego, en cada atractor. No obstante, esta caracterización no ofrece detalle alguno respecto del tipo de interacciones no lineales que llevan a la estructuración topológica de un atractor en el espacio de fases.

La topología de un atractor de un sistema dinámico determinístico en el espacio de fases es reconocida como un invariante en un sistema caótico determinístico. ¿Es posible entonces describir la topología de estos atractores utilizando reconstrucciones parciales del espacio de fases, como se ha hecho anteriormente para otros sistemas? Esta es la pregunta que motiva esta tesis, en la cual hemos intentado utilizar un tipo particular de cálculo topológico basado en los conceptos de persistencia homológica, para caracterizar estas diferencias.

La metodología de homologías persistentes ha sido utilizada con éxito para clasificar proteínas con estructuras diferenciadas, o para la detección en tiempo real de crisis epilépticas. En el ámbito de las ciencias del clima, existen precedentes que utilizan estas técnicas para detectar ríos atmosféricos.

Utilizando estrategias inspiradas en el estudio de la variabilidad climática en otros escenarios, hemos logrado extraer una serie de modos estadísticamente significativos aplicando métodos multicanal de análisis de espectros singulares. Estos resultados permiten

generar nubes de puntos que, en tres dimensiones, condensan los rasgos principales de la evolución caótica de las variables atmosféricas u oceánicas para los distintos atractores, en ausencia del ciclo estacional, que es voluntariamente removido.

Se han construido diagramas de persistencia para los distintos atractores, y se ha establecido un listado de marcadores topológicos, similares a los utilizados en los estudios de otras disciplinas. Calculando el valor de los marcadores topológicos para los cuatro atractores considerados, se encuentra que cada uno de ellos tiene un rasgo predominante que lo distingue de los otros en ventanas temporales de mil días.

Los métodos de homología persistente no permiten dilucidar con claridad la estructura global del atractor en términos de una variedad enramada. Los diagramas de persistencia nos ofrecen una caracterización topológica menos precisa, que no es lo suficientemente exhaustiva como para identificar el tipo de atractor en cada caso. No obstante, nuestros resultados marcan un camino hacia la construcción de complejos celulares representativos de la estructura topológica de cada atractor climático, utilizando los marcadores como guía.

Nuestros resultados son relativos al pre-procesamiento realizado, a la ventana temporal seleccionada, a las componentes principales retenidas, y también, por supuesto, a los marcadores topológicos seleccionados. No son estos los únicos que pueden construirse. No obstante, las dinámicas parecen estar lo suficientemente diferenciadas como para separar satisfactoriamente los atractores Hotstate, Warmstate, Coldstate y Waterbelt, considerados en este trabajo.

Nuestros esfuerzos constituyen un primer mojón hacia una caracterización topológica de flujos asociados a atractores climáticos en el espacio de fases. La perspectiva de construir un complejo, e incluso un ‘templex’ (Charó et al., 2022), es decir, un complejo dotado de un grafo con la información del flujo entre celdas, que resulte consistente con las propiedades identificadas para cada uno de estos atractores, se abre a partir de este primer esbozo de las características básicas que debería presentar una estructura topológica capaz de diferenciar un atractor climático de otro.

# Agradecimientos

A Gisela, Denisse, Pablo, Maura.

Al DM, al DCAO, al CIMA, a IFAECI, a la Universidad de Ginebra y a la FCEN.

A mis amigos.

A mi familia.



# Bibliografía

- Bauer, U. (2021). Ripser: efficient computation of vietoris–rips persistence barcodes. *Journal of Applied and Computational Topology*, 5(3):391–423.
- Broomhead, D. S. and King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2-3):217–236.
- Brunetti, M., Kasparian, J., and V erard, C. (2019). Co-existing climate attractors in a coupled aquaplanet. *Climate Dynamics*, 53:6293–6308.
- Cang, Z., Mu, L., Wu, K., Opron, K., Xia, K., and Wei, G.-W. (2015). A topological approach for protein classification. *Computational and Mathematical Biophysics*, 3(1).
- Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308.
- Carlsson, G. (2014). Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289–368.
- Carlsson, G. and Zomorodian, A. (2007). The theory of multidimensional persistence. In *Proceedings of the twenty-third annual symposium on Computational geometry*, pages 184–193.
- Char o, G. D., Chekroun, M. D., Sciamarella, D., and Ghil, M. (2021). Noise-driven topological changes in chaotic dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(10):103115.
- Char o, G. D., Letellier, C., and Sciamarella, D. (2022). Templex: A bridge between homologies and templates for chaotic attractors. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(8):083108.
- Char o, G. D., Ghil, M., and Sciamarella, D. (2022). Random templex encodes topological tipping points in noise-driven chaotic dynamics.
- Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963.
- Dettinger, M. D., Ghil, M., and Keppenne, C. L. (1995). Interannual and interdecadal variability in united states surface-air temperatures, 1910-87. *Climatic Change*, 31(1):35–66.
- D ıaz S anchez, O. A. et al. (2019). Fundamentos de la homolog ıa persistente y su c odigo de barras asociado.
- Edelsbrunner, H. and Harer, J. (2008). Persistent homology-a survey. *Contemporary mathematics*, 453:257–282.
- Edelsbrunner, H. and Harer, J. L. (2022). *Computational topology: an introduction*. American Mathematical Society.

- Falasca, F. and Bracco, A. (2022). Exploring the tropical pacific manifold in models and observations. *Physical Review X*, 12(2):021054.
- Faranda, D., Alvarez-Castro, M. C., Messori, G., Rodrigues, D., and Yiou, P. (2019). The hamman effect or how a warm ocean enhances large scale atmospheric predictability. *Nature communications*, 10(1):1316.
- Faranda, D., Messori, G., and Yiou, P. (2017). Dynamical proxies of north atlantic predictability and extremes. *Scientific reports*, 7(1):1–10.
- Feliks, Y., Ghil, M., and Robertson, A. W. (2010). Oscillatory climate modes in the eastern mediterranean and their synchronization with the north atlantic oscillation. *Journal of Climate*, 23(15):4060–4079.
- Fernández, X., Borghini, E., Mindlin, G., and Groisman, P. (2023). Intrinsic persistent homology via density-based metric learning. *Journal of Machine Learning Research*, 24(75):1–42.
- Ferreira, D., Marshall, J., Ito, T., and McGee, D. (2018). Linking glacial-interglacial states to multiple equilibria of climate. *Geophysical Research Letters*, 45(17):9160–9170.
- Fraser, A. M. and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2):1134.
- Ghil, M. (1976). Climate stability for a sellers-type model. *Journal of Atmospheric Sciences*, 33(1):3–20.
- Ghil, M., Allen, M., Dettinger, M., Ide, K., Kondrashov, D., Mann, M., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., et al. (2002). Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40(1):3–1.
- Ghil, M. and Sciamarella, D. (2023). Dynamical systems, algebraic topology, and the climate sciences. *EGUsphere*, pages 1–53.
- Ghrist, R. (2008). Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75.
- Gilmore, R. (1998). Topological analysis of chaotic dynamical systems. *Reviews of Modern Physics*, 70(4):1455.
- Gilmore, R. and Lefranc, M. (2002). The topology of chaos: Alice in stretch and squeeze land. *Hoboken: Wiley & sons inc*, page 518.
- Groth, A., Feliks, Y., Kondrashov, D., and Ghil, M. (2017). Interannual variability in the north atlantic ocean’s temperature field and its association with the wind stress forcing. *Journal of Climate*, 30(7):2655–2678.
- Groth, A. and Ghil, M. (2011). Multivariate singular spectrum analysis and the road to phase synchronization. *Physical Review E*, 84(3):036206.
- Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202.
- Kennel, M. B., Brown, R., and Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6):3403.



- Lucarini, V. and Bódai, T. (2017). Edge states in the climate system: exploring global instabilities and critical transitions. *Nonlinearity*, 30(7):R32.
- Lucarini, V., Faranda, D., de Freitas, J. M. M., Holland, M., Kuna, T., Nicol, M., Todd, M., Vaienti, S., et al. (2016). *Extremes and recurrence in dynamical systems*. John Wiley & Sons.
- Maletić, S., Zhao, Y., and Rajković, M. (2016). Persistent topological features of dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(5):053105.
- Mindlin, G. (2017). Dinámica no lineal. *Universidad Nacional de Quilmes: Bernal, Argentina*.
- Moron, V., Vautard, R., and Ghil, M. (1998). Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dynamics*, 14:545–569.
- Muldoon, M. R., MacKay, R. S., Huke, J. P., and Broomhead, D. S. (1993). Topology from time series. *Physica D: Nonlinear Phenomena*, 65(1-2):1–16.
- Munkres, J. R. (2018). *Elements of algebraic topology*. CRC press.
- Murugan, J. and Robertson, D. (2019). An introduction to topological data analysis for physicists: From lgm to frbs. *arXiv preprint arXiv:1904.11044*.
- Natiello, M. A. et al. (2007). *The user's approach to topological methods in 3d dynamical systems*. World Scientific.
- Perea, J. A. (2019). Topological time series analysis. *Not Am Math Soc*, 66(5):686–694.
- Perea, J. A. and Harer, J. (2015). Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15:799–838.
- Ragon, C., Lembo, V., Lucarini, V., V erard, C., Kasparian, J., and Brunetti, M. (2022). Robustness of competing climatic states. *Journal of Climate*, 35(9):2769–2784.
- Sandri, M. (2003). Melissa: un package per mathematica per l'analisi singular-spectrum delle serie temporali. guida per l'utente.
- Sciamarella, D. and Mindlin, G. B. (1999). Topological structure of chaotic flows from human speech data. *Physical Review Letters*, 82(7):1450.
- Sciamarella, D. and Mindlin, G. B. (2001). Unveiling the topological structure of chaotic flows from data. *Physical Review E*, 64(3):036209.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.
- Smale, S. (1980). The mathematics of time. *Springer-Verlag*.
- Steffen, W., Rockstr om, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., Summerhayes, C. P., Barnosky, A. D., Cornell, S. E., Crucifix, M., et al. (2018). Trajectories of the earth system in the anthropocene. *Proceedings of the National Academy of Sciences*, 115(33):8252–8259.
- Tang, Y., Li, L., Dong, W., and Wang, B. (2016). Reducing the climate shift in a new coupled model. *Science Bulletin*, 61:488–494.

- Tralie, C., Saul, N., and Bar-On, R. (2018). Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925.
- Tuerke, F., Pastur, L., Fraigneau, Y., Sciamarella, D., Lusseyran, F., and Artana, G. (2017). Nonlinear dynamics and hydrodynamic feedback in two-dimensional double cavity flow. *Journal of Fluid Mechanics*, 813:1–22.
- Vautard, R. and Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D: Nonlinear Phenomena*, 35(3):395–424.
- Wadhwa, R. R., Dhawan, A., Williamson, D. F., and Scott, J. G. (2018). A flat persistence diagram for improved visualization of persistent homology. *arXiv preprint arXiv:1812.04567*.
- Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532.
- Zomorodian, A. and Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274.