



Flujo de Información en Sistemas Dinámicos Extendidos

Trabajo Especial de Licenciatura en Física

Facultad de Matemática, Astronomía, Física y Computación
Universidad Nacional de Córdoba

Santiago Rosa

Directores: Manuel Pulido y Pedro Pury



Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional.

Quiero agradecer en primer lugar a mi mamá, Mariana, y a mi hermana, Sofía, por su amor incondicional, y a mi tía, Graciela, por su cariño y apoyo estos últimos años. A Manuel, mi director, por todas las horas de videollamada y cadenas de mails y por introducirme en este tema de investigación. A Pedro, mi otro director, por sacarme muchas dudas sobre redes neuronales y guiarme con el escrito. Al tribunal, Guille, Negui y Pancho, por tomarse el tiempo de leer y evaluar este trabajo. A Marcos, Juan y Mauri, con quienes estudié y rendí la gran mayoría de las materias de la carrera. A todes mis amigues, con quienes compartí hermosos momentos todos estos años. Por último, al Estado y al pueblo Argentino, quienes apoyan y sustentan la universidad nacional, pública y gratuita.

Resumen

En este trabajo se evalúan cuantificadores de información, particularmente la información mutua y la transferencia de entropía, como medidas de correlación entre diferentes variables de sistemas que simulan procesos en la atmósfera. La dinámica atmosférica podría caracterizarse, en esencia, como un sistema no lineal dinámico extendido espacialmente con complejas interacciones entre las diferentes escalas que componen su dinámica, como también entre los distintos procesos físicos involucrados en ella. Para evaluar si las medidas de información pueden capturar y cuantificar estas interacciones, se aplican los cuantificadores de información al modelo Lorenz-96, un sistema dinámico no lineal en el cual se incluyen interacciones de largo alcance entre escalas; mientras que para detectar cascadas de energía, se utiliza un modelo sintético de turbulencia 3D, y la ecuación barotrópica de la vorticidad. Para estimar las distribuciones de probabilidad y los cuantificadores de información se analizan las series temporales de los procesos de interés a través del análisis simbólico ordinal usando el método de Bandt-Pompe de reducción simbólica de datos en las señales, vía aproximación por primeros vecinos usando el método de Kraskov-Stögbauer-Grassberger, y entrenando una red neuronal utilizando el estimador neuronal de información mutua. Comparando diferentes experimentos, se muestra que incluso cuando la dinámica de las variables de la gran escala de los sistemas tienen correlaciones de corto alcance, las variables de la pequeña escala pueden dar lugar a flujos de información de largo alcance, y que las medidas de información son herramientas útiles para establecer el flujo de información observacional en sistemas dinámicos, así como también la detección parcial de cascadas de energía en modelos de turbulencia.

Abstract

In this work Information quantifiers are evaluated, particularly mutual information and transfer entropy, as correlation measurements between different variables of systems simulating processes in the atmosphere. The dynamic of the Atmosphere could be characterized, in essence, as a spatially extended, non-linear dynamical system with complex interactions between the different scales that composes its dynamics, as well as between the different physical processes involved in it. To evaluate if the information measurements can capture and quantify these interactions, these quantifiers are applied to the Lorenz-96 model, a non-linear dynamical system in wich long range interactions between scales are included; while to detect energy cascades, a sintetical 3D turbulence model is used, as well as the barotropic vorticity equation. In order to estimate the probability distributions and information quantifiers, the time series of the processes of interest are analyzed through symbolic ordinal analysis using the Bandt and Pompe method of symbolic data reduction in signals, via first neighbours approach using the Kraskov-Stögbauer-Grassberger method, and training a neural network using the mutual information neural estimator. Comparing different experiments, it is shown thet even when the dynamics of the variables in the systems grand scales have short range correlations, the small scale variables can lead to long range information flux, and that the information measurements are useful tools to establish observational information flux in dynamical systems, as well to detect partial energy cascades in turbulence models.

Clasificación:

- 02.70.-c Computational techniques; simulations
- 05.45.Tp Time series analysis
- 89.70.+c Information theory and communication theory
- 92.60.hk Convection, turbulence, and diffusion

Palabras clave: Teoría de información, atmósfera, correlación, cascadas de energía

Índice general

1	Introducción	7
1.1	Teoría de información de Shannon	9
1.2	Análisis simbólico Ordinal (ASO)	12
1.3	Método de Kraskov, Stögbauer y Grassberger (KSG)	15
1.4	Estimación de información mutua mediante redes neuronales (MINE)	17
2	Sistemas dinámicos	22
2.1	Modelo Lorenz-96	22
2.2	Modelo de Gledzer-Ohkitana-Yamada (GOY)	26
2.3	Ecuación barotrópica de la vorticidad	28
3	Experimentos con el Lorenz 96	33
3.1	Análisis simbólico ordinal	33
3.2	Método de KSG	36
3.3	Redes neuronales	38
3.4	Comparación entre estimadores	41
4	Experimentos con el GOY	43
4.1	Análisis simbólico Ordinal	43
4.2	Método de KSG	44
4.3	Redes Neuronales	46
4.4	Comparación entre estimadores	48
5	Experimentos con la ecuación barotrópica de la vorticidad	50
5.1	Análisis simbólico ordinal	50
5.2	Método de KSG	53
6	Conclusiones	55
	Bibliografía	55

1 Introducción

La atmósfera es un sistema multiescala complejo, con interacciones entre los múltiples procesos físicos que la involucran. Cualquier evento meteorológico, sea severo o estadísticamente probable, es consecuencia del acoplamiento de varios subsistemas que la componen. Pero además la atmósfera está fuertemente acoplada a otros grandes sistemas, tales como el océano y su dinámica, el suelo y sus procesos, el ecosistema, etc. los cuales también cumplen un rol preponderante en la determinación de un evento meteorológico en consideración [1]. El estudio de estos eventos meteorológicos en gran parte se centra en determinar las causas que lo ocasionaron, es decir, en conectar el evento en consideración con alguna característica de los subsistemas que pueda desencadenar el evento.

El sistema climático puede ser pensado como una red. Los nodos pueden representar variables en puntos de una red, o modos dinámicos. Está demostrado que la arquitectura de la red climática es del tipo mundo pequeño [2]: por ejemplo, la región tropical es una subred que está completamente conectada entre los nodos, mientras que la dinámica de medias latitudes es una red de escala libre que está caracterizada por super nodos dominantes [3].

El concepto de redes complejas y la teoría desarrolladas para su tratamiento puede ser una herramienta útil para entender el complejo comportamiento de la atmósfera y sus variables. En este sentido en la dinámica atmosférica es de vital importancia entender las interacciones entre los nodos, así como también la relación causa-efecto entre los nodos o procesos que constituyen el sistema. Una forma posible de determinar causalidad en los modos es mediante el *test de causalidad de Granger* [4], el cual determina si una serie temporal es útil para pronosticar otra variable. Sean tres variables X_t , Y_t y W_t a tiempo t , y se quiere predecir X_{t+1} usando como información a los términos pasados de X_t y W_t . También, de forma independiente, se predice X_{t+1} usando X_t , W_t y adicionalmente términos pasados de Y_t . Si el segundo pronóstico de X_{t+1} es más exitoso de acuerdo a ciertas funciones de costo, entonces el pasado de Y contiene información importante para predecir X_{t+1} que no está en el pasado de X_t ó W_t . Entonces, Y_t será *Granger-causal* (G -causal) de X_{t+1} si: (a) Y_t ocurre antes que X_{t+1} ; y (b) si contiene información útil para pronosticar X_{t+1} que no es encontrada en otro grupo de variables apropiadas.

Esta medida usualmente es aplicada en el contexto de modelos autorregresivos lineales. Dado el carácter caótico de la atmósfera, se busca una medida que logre captar correlaciones en series temporales altamente no lineales.

Otra forma (más simple) de detectar correlaciones es calcular el coeficiente de correlación de Pearson entre las series temporales de las variables X_i , X_j de la red. La desventaja de este método es que mide qué tan *linealmente correlacionadas* están las variables. La no linealidad de la evolución de las variables, así como la no gaussianidad

de las distribuciones de probabilidad asociadas a dichas variables hacen que el coeficiente de Pearson no sea una buena medida de correlación. Esto es de particular importancia para la dinámica atmosférica donde se tiene una evolución altamente no lineal que desencadena distribuciones no gaussianas e incluso multimodales, esto ha sido encontrado aún para tiempos de análisis relativamente cortos (seis horas).

Para ilustrar la no linealidad y no gaussianidad de los procesos atmosféricos, consideremos la evolución temporal del sistema dinámico *Lorenz 63*, o más comúnmente denominado *atractor de Lorenz* [5]. Este sistema de ecuaciones diferenciales no lineales ordinarias rige la evolución de una capa bidimensional de fluido calentada uniformemente desde la base y enfriada uniformemente desde arriba. En este sistema, variaciones infinitesimales en las condiciones iniciales producen perturbaciones que crecen exponencialmente en el tiempo en el espacio de fases del sistema, a tal punto de hacer imposible la predicción del sistema en un tiempo futuro. En la figura 1.1 se muestra una distribución gaussiana aleatoria de puntos a los cuales se los hace evolucionar en el sistema, y al cabo de un tiempo la distribución es altamente no gaussiana.

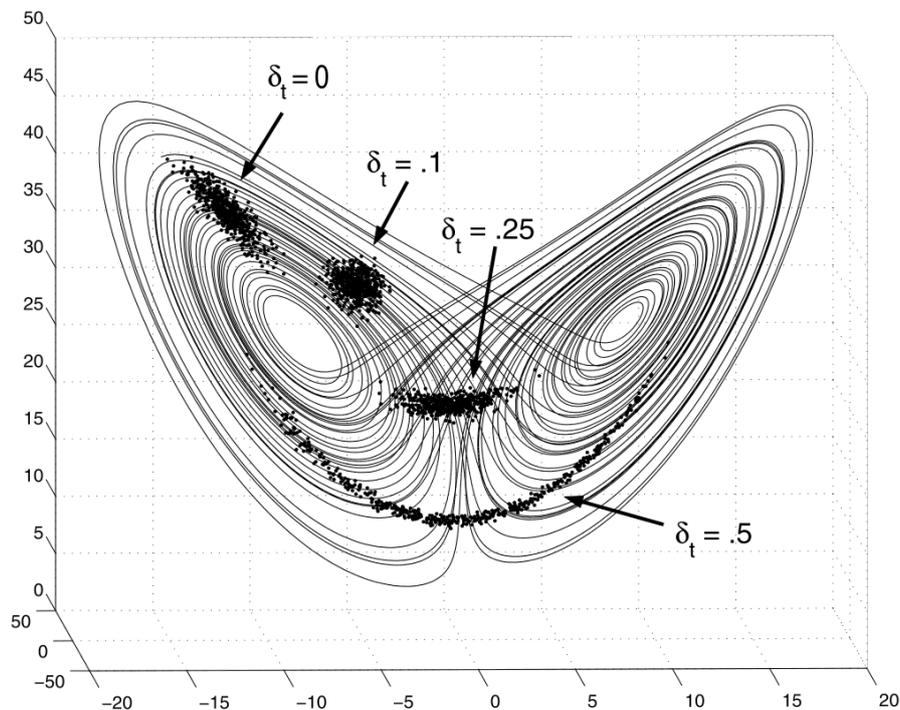


Figura 1.1: Distribución inicial gaussiana evolucionando en el atractor de Lorenz.

Este simple modelo de convección atmosférica muestra que tomar como función de correlación de variables al coeficiente de Pearson o el test de causalidad de Granger es ineficiente, y al hacerlo se perderían todo tipo de interacciones no lineales que no son cuantificadas por esas medidas de causalidad.

La alternativa que se propone en este trabajo es el uso de *cuantificadores de información clásica* como medidas de correlación no lineal, en donde se trabaja directamente con

las distribuciones de probabilidad de las series temporales sin hacer suposiciones acerca de las distribuciones ni del tipo de correlación entre las series a comparar. La teoría de la información, propuesta en 1948 por Claude Shannon para encontrar los límites fundamentales en el procesamiento de señales y operaciones comunicacionales como la compresión de datos [6], hoy en día es aplicada a diversos campos como la matemática, ciencias de la computación, física, neurobiología, y varios otros.

1.1. Teoría de información de Shannon

Shannon buscaba una medida de información que satisfaga las condiciones:

1. Debe ser aditiva para piezas independientes de información,
2. Debe reflejar probabilidad de eventos, en particular, capturar el aumento de incertidumbre asociada al aumento del número de eventos igualmente probables,
3. Debe ser continua con respecto a los cambios en estas probabilidades.

Consideremos muestras de una variable aleatoria X de un evento x , que puede tomar valores de un espacio muestral Ω_X . Si $p(x)$ es la probabilidad de ocurrencia del evento x se define la *información de un evento x* como

$$\eta(x) = -\log_2 p(x). \quad (1.1)$$

Esta magnitud puede variar a lo largo del tiempo, o sobre varias realizaciones del sistema, siendo los dos casos equivalentes si el sistema considerado es ergódico. Si se promedia la información sobre todos los eventos posibles, se obtiene la *Entropía de Shannon*:

$$\mathbf{H}(X) = E\{\eta(x)\} = - \sum_{x \in \Omega_X} p_X(x) \log_2(p_X(x)) \quad (1.2)$$

La interpretación de esta definición es la siguiente: La entropía $\mathbf{H}(X)$ de una variable aleatoria X es una medida de la cantidad de incertidumbre asociada a la variable X . H es mínima y toma el valor $\mathbf{H}(X) = 0$ cuando el sistema está en un único estado x_i , mientras que es máxima y toma el valor $\mathbf{H}(X) = M$ cuando todos los M estados x_i , $i = 1, \dots, M$ tienen igual probabilidad de ocurrencia $1/M$. La base 2 del logaritmo implica que la entropía está dada en *bits*. En otras palabras, H es el número óptimo promedio de bits necesarios para cifrar realizaciones de la variable aleatoria X .

En este trabajo se utilizará la base e , por lo que el resultado estará dado en *bits naturales*, y se denotará de ahora en más por $\log(x)$ al logaritmo natural de x .

La definición 1.2 es extendible a más de una variable: Se define la entropía conjunta entre las variables aleatorias X e Y como

$$\mathbf{H}(X, Y) = - \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X} p_{X,Y}(x, y) \log(p_{X,Y}(x, y)) \quad (1.3)$$

donde $p(x, y)$ es la probabilidad conjunta de los eventos x e y , o sea, que simultáneamente se den x e y .

Observar una variable X condicionada por otra variable Y siempre es mejor que observar sólo la variable X . En este sentido se define la entropía condicionada como

$$\mathbf{H}(X|Y) = - \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(y)} \quad (1.4)$$

Para construir un cifrado óptimo que use la cantidad justa cuantificada por la entropía de Shannon, es necesario conocer exactapente la distribución de probabilidad $p_X(x)$. El exeso de bits cuantificados por la entropía si se utiliza una aproximación $q_X(x)$ en vez de $p_X(x)$ estará dado por la *divergencia de Kullback-Leibler* [7]:

$$D_{KL}(p||q) = \sum_{x \in \Omega_X} p_X(x) \log \left(\frac{p_X(x)}{q_X(x)} \right) \quad (1.5)$$

Finalmente, se define la *información mutua* entre las variables aleatorias X e Y como

$$\mathbf{I}(X, Y) = - \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X} p_{X,Y}(x, y) \log \left(\frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} \right) \quad (1.6)$$

La información mutua es la cantidad de información compartida entre X e Y . Es una medida de su dependencia estadística, y puede ser pensada como una forma de correlación no lineal: $\mathbf{I}(X, Y) = 0 \iff X$ e Y son independientes, es decir, cuando $p_{X,Y}(x, y) = p_X(x)p_Y(y)$. La información mutua tiene las siguientes propiedades:

1. Es no negativa: $I(X, Y) \geq 0$,
2. Es simétrica: $I(X, Y) = I(Y, X)$,
3. Usando el Teorema de Bayes, puede ser expresada en función de las entropías conjunta y marginales:

$$\mathbf{I} = H(X) + H(Y) - H(X, Y), \quad (1.7)$$

considerando la ecuación 1.5, se puede reescribir la información mutua:

$$I(X, Y) = D_{KL}(p_{(X,Y)} || p_X p_Y)$$

Esta identidad va a ser utilizada más adelante en este trabajo.

Finalmente, si se condiciona la información mutua entre X e Y , con otra variable Z , se obtiene la *información mutua condicionada*:

$$\mathbf{I}(X, Y|Z) = - \sum_{z \in \Omega_Z} p_Z(z) \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X} p_{X,Y|Z}(x, y|z) \log \left(\frac{p_{X,Y|Z}(x, y|z)}{p_{X|Z}(x|z)p_{Y|Z}(y|z)} \right) \quad (1.8)$$

que también se puede escribir en función de las entropías conjuntas y condicionadas:

$$\mathbf{I}(X : Y|Z) = \mathbf{H}(X, Z) + \mathbf{H}(Y, Z) - \mathbf{H}(X, Y, Z) - \mathbf{H}(Z) \quad (1.9)$$

Todo el análisis hasta ahora se realizó bajo la suposición de que X e Y son variables aleatorias estáticas. Supongamos que ahora se tiene una secuencia de realizaciones conjuntas (x_t, y_t) de X e Y enumeradas secuencialmente por el índice temporal $t = 1, 2, 3, \dots$. El proceso (X_t, Y_t) puede ser pensado como la evolución temporal de variables aleatorias X, Y de ciertos procesos estocásticos (precios de la bolsa de valores, señales neuronales, etc) ó de sistemas dinámicos deterministas.

Se podría considerar la información mutua a cierto tiempo $\mathbf{I}(X_t, Y_t)$, pero cuando nos referimos a *secuencia de realizaciones conjuntas a tiempo t* , podría haber dependencias entre subconjuntos $\{X_t, Y_s, t \in T, s \in S\}$ de las variables individuales, es decir, variables a tiempo t pueden depender de otras en un tiempo pasado s . [8] y [9], de forma independiente, vieron que la información mutua atrasada $\mathbf{I}(X_t, Y_{t-\tau})$ no es una medida satisfactoria dado que falla a la hora de tener en cuenta la historia en común de los procesos X e Y : para evaluar la influencia del pasado de Y con el presente de X , la información compartida entre X y su propio pasado debe ser tenida en cuenta.

Esto motiva a la definición de la *transferencia de entropía* (TE):

$$\mathbf{T}_{Y \rightarrow X}(t) \equiv \mathbf{I}(X_t, Y_{t-\tau}, X_{t-\tau}) \quad (1.10)$$

$\mathbf{T}_{Y \rightarrow X}(t)$, al ser una información mutua condicionada, es siempre no-negativa, y es una forma de dependencia estadística entre X e Y : $\mathbf{T}_{Y \rightarrow X} = 0 \iff X$, condicionada con su propio pasado, es independiente del pasado de Y .

La definición de transferencia de entropía puede ser extendida a más variables, obteniendo estimaciones más precisas de información mutua mientras más variables del sistema se conocen. El problema que surge inmediatamente es el aumento de la dimensionalidad de las distribuciones de probabilidad (problema comunmente denominado *curso de la dimensionalidad*), aumentando de forma exponencial la cantidad de muestras requeridas de las PDF, y por lo tanto el tiempo de cómputo por cada dimensión que se agrega. Una alternativa posible al problema de la dimensionalidad es embeber la TE en el marco de modelos gráficos [10], en el cual la TE se descompone en una suma de contribuciones de dimensión finita [11]. Los autores utilizan por un lado series temporales de un modelo no lineal estocástico, y por otro observaciones de anomalías de presión a nivel del mar en los meses de invierno de 1997-2003 en cuatro localizaciones de Europa del Este.

Otro uso de la TE es la detección de *cascaídas de energía* en modelos sintéticos de turbulencia [12], los cuales tienen un espectro de la energía en función del número de onda que siguen la *ley de similitud de Kolmogorov* $E(k) \propto k^{-5/3}$. Se utiliza una forma de TE normalizada y bivariada, estimando la distribuciones de probabilidad con histogramas, para detectar transferencias de energía desde modos de energía con bajo número de onda hacia los modos con alto número de onda, calculando las magnitudes de las interacciones, así también como los tiempos característicos en los cuales ocurren [13].

Medidas de información se utilizaron también para la correcta selección de parámetros en parametrizaciones de subgrilla [14] en sistemas dinámicos que simulan comportamientos de ciertos procesos atmosféricos [15]. Los autores compararon series temporales observadas de un *modelo verdadero* con cierta dinámica oculta, y lo estimaron con un modelo más simple con parametrizaciones determinísticas y estocásticas. Para obtener el mejor conjunto de parámetros, se compararon las series temporales del modelo verdadero con las de los modelos parametrizados utilizando como medidas a la *complejidad estadística* [16] y a la *Divergencia de Jensen-Shannon* [17]. En este trabajo utilizaron medidas de información que requieren distribuciones de probabilidad de dimensión 1, ya que no utilizaron nunca distribuciones de probabilidad conjunta.

Este trabajo sigue la metodología de [13]: utilizando tres formas de estimar la información mutua (Análisis simbólico ordinal, búsqueda de primeros vecinos y redes neuronales) y una forma de estimar la transferencia de entropía (análisis simbólico ordinal), el objetivo es aplicar las medidas ya mencionadas para:

1. cuantificar las interacciones entre variables y entre escalas en sistemas dinámicos extendidos que imitan procesos atmosféricos (Modelo Lorenz 96 de una y dos escalas), como también detectar la direccionalidad de estas interacciones;
2. Detectar transferencias de energía desde los números de onda bajos hacia los números de onda altos en un modelo sintético de turbulencia en tres dimensiones (modelo de Gledzer-Ohkitana-Yamada);
3. Detectar transferencias de enstrofia desde los números de onda bajos hacia los números de onda altos en un modelo realista de turbulencia bidimensional (ecuación barotrópica de la vorticidad).

En las siguientes secciones se describirán los métodos alternativos para estimar la distribución de probabilidad de una serie temporal (*Análisis simbólico Ordinal*), y para estimar la información mutua (Estimación por k primeros vecinos, y mediante el uso de redes neuronales).

1.2. Análisis simbólico Ordinal (ASO)

El cálculo de cuantificadores de información en sistemas dinámicos clásicos requiere de antemano el conocimiento de la distribución de probabilidad asociada a las series temporales que se quieren analizar. El método más usado es el histograma: El espacio de los posibles estados es particionado en bins, y contando el número de veces N_i que las trayectorias de un ensamble pasan por el i -ésimo bin en un dado tiempo, se define la probabilidad $p_i = N_i/N$, donde N es el número total de trayectorias posibles del sistema. La desventaja del método es que la PDF resultante es muy dependiente del ancho de bins utilizado.

Una forma alternativa de estimar la PDF es mediante secuencias temporales. Supongamos una secuencia de L tiempos y etiquetamos los bins, por lo que podemos formar

una secuencia simbólica de L símbolos, dado que en toda la serie temporal el sistema recorre L bins. En esta secuencia, cada símbolo proveniente de un conjunto finito representa un bin, y un patrón se forma con una secuencia finita de bins, que es recorrida por la trayectoria en L pasos de tiempo. Contando la ocurrencia de cada patrón sobre el total de secuencias, determinamos la distribución de probabilidad. Si disminuimos el tamaño de los bins hasta tenderlos a cero, se deriva de esta probabilidad la entropía de Kolmogorov-Sinai [18].

[19] propusieron una metodología simple y robusta para la estimación de la distribución de probabilidad asociada a una serie temporal $\{x_t\}$, basada en una secuencia simbólica que surge naturalmente de la serie, sin suposición alguna sobre el modelo [19]. Nos referimos a *análisis simbólico ordinal* a esta metodología. El método consiste en traducir la serie temporal a una serie de patrones ordinales de dimensión D (llamémoslos π), los cuales son creados de la siguiente manera: (i) se asigna a cada tiempo t el vector D -dimensional cuyas componentes son los valores de la serie temporal hasta el tiempo t :

$$\sigma_t \rightarrow (x_{t-(D-1)}, x_{t-(D-2)}, \dots, x_{t-1}, x_t) \quad (1.11)$$

(ii) se reordenan las componentes de σ_t en orden ascendente y se aplica la misma permutación al vector ordinal $\pi_0 = (0, 1, \dots, D-1)$. El *patrón ordinal a tiempo* (t) es, finalmente, la permutación $\pi = (r_0, r_1, \dots, r_{D-1})$ de π_0 definida por

$$x_{t-r_{D-1}} \leq x_{t-r_{D-2}} \leq \dots \leq x_{t-r_1} \leq x_{t-r_0} \quad (1.12)$$

De esta forma, el vector (1.11) es convertido en un único símbolo π . Definimos $r_i \leq r_{i-1}$ si $x_{s-r_i} = x_{s-r_{i-1}}$ para tener unicidad.

La ocurrencia de cada patrón se cuenta en toda la serie temporal, y la probabilidad de aparición de cada símbolo π es el número de veces que aparece el patrón dividido el número total de patrones de la serie temporal. Entonces, la función distribución de probabilidad de Bandt y Pompe (BP-PDF) está dada por $P = \{p(\pi_i), i = 1, \dots, D!\}$, con

$$p(\pi_i) = \frac{\#\{s | t \leq M - (D-1); (t) \text{ de tipo } \pi_i\}}{M - (D-1)} \quad (1.13)$$

donde $\#$ denota la cardinalidad, y M es el largo de la serie temporal.

Para ilustrar el análisis simbólico ordinal, consideremos un ejemplo: tenemos una serie temporal con siete ($M = 7$) valores: $\{4, 7, 9, 10, 6, 11, 3\}$ y queremos calcular la BP-PDF con $D = 3$. En este caso el espacio de estados está dividido en $3!$ particiones, por lo que tengo 6 patrones de permutación mutuamente excluyentes. Los tripletes $(4, 7, 9)$ y $(7, 9, 10)$ representan el patrón de permutación $\{012\}$ dado que ya están en orden ascendente. Por otro lado, $(9, 10, 6)$ y $(6, 11, 3)$ corresponden al patrón de permutación $\{201\}$ dado que $x_{t+2} < x_t < x_{t+1}$, mientras que $(10, 6, 11)$ tiene el patrón de permutación $\{102\}$. Entonces las probabilidades asociadas a estos 6 patrones son: $p(\{012\}) = p(\{201\}) = 2/5$; $p(\{102\}) = 1/5$; $p(\{021\}) = p(\{210\}) = 0$.

Ahora que se tiene una estimación de la distribución de probabilidad de una serie temporal, es sencillo definir la *entropía de permutación*:

$$H(n) = - \sum_{\pi} p(\pi) \log p(\pi) \quad (1.14)$$

donde la suma corre sobre todas las $n!$ permutaciones π de orden D . Claramente se ve que $0 \leq H(n) \leq \log n!$, donde el límite inferior corresponde a series enteramente crecientes, decrecientes o constantes, y el límite superior a series completamente aleatorias.

Para calcular la distribución de probabilidad conjunta, se cuenta la cantidad de veces que el conjunto de símbolos (π_i, π_j) , con $i, j = 1, 2, \dots, D!$ aparece a lo largo de toda la serie bivariada $(\{x\}_t, \{y\}_t)$, y dividiendo por la cantidad de patrones que hay en la serie temporal $(N - (D - 1))$, si N es el largo de la serie). Entonces, la información mutua es

$$\mathbf{I}(X, Y) = - \sum_{\pi_y} \sum_{\pi_x} p(\pi_x, \pi_y) \log \left(\frac{p(\pi_x, \pi_y)}{p(\pi_x)p(\pi_y)} \right) \quad (1.15)$$

Para estimar la información mutua entre una variable a tiempo presente y otra a tiempo pasado τ , donde τ se mide en tiempos de observación t_{obs} de las series temporales, se toman los últimos $N - \tau$ valores $\{x_\tau, x_{\tau+1}, \dots, x_N\}$ de la serie temporal presente y los $N - \tau$ primeros valores $\{y_0, y_1, \dots, y_{N-\tau}\}$, es decir, se atrasa en un tiempo τ la serie pasada.

Utilizando la identidad 1.9, atrasando correctamente las variables y calculando todas las distribuciones conjuntas requeridas, se calcula de la misma forma la transferencia de entropía.

Cuando se traducen las series temporales a series de símbolos, en principio cada símbolo es un vector D -dimensional, y comparar vectores término a término para contarlos y obtener la PDF se hace computacionalmente demandante cuando la serie temporal es significativamente larga ($\sim 10^5$ puntos).

Hay $D!$ formas de permutar D objetos, las cuales pueden ser representadas por un número entero entre 0 y $D! - 1$. Una forma de lograr esto es utilizando el *código de Lehmer*, el cual mapea de forma biyectiva permutaciones a decimales. En el algoritmo 2 se presenta el método mediante un ejemplo. Así, se transcriben las series temporales a números enteros (símbolos). Contando la frecuencia de aparición de todos los números en la serie se obtiene la función distribución de probabilidad de Bandt y Pompe. Tanto el código de Lehmer como la obtención de la PDF y la cuantificación de las medidas de información mutua y transferencia de entropía fueron implementadas en Fortran 90.

Algoritmo 2 Código de Lehmer

Paso 1. Sea $i = \arg(\min(\pi))$ **3,4,1,2**

Paso 2. Contar en número de dígitos a la izquierda del dígito i que sean mayores que $\pi(i)$. Este es el primer dígito del factorádico.

3,4,1,2 número de dígitos: **2**

Paso 3. Remover $\pi(i)$ de la permutación. **3,4,2**

Paso 4. Repetir los pasos 1-3 hasta que no queden dígitos:

3,4,2 \rightarrow **3,4,2** número de dígitos: **2**

3,4 \rightarrow **3,4** número de dígitos: **0**

4 \rightarrow **4** número de dígitos: **0**

Paso 5. convertir los factorádicos **2,2,0,0** a decimal, usando el factorial de la posición de $\pi(i)$ como base para cada factorádico:

lugar:	3!	2!	1!	0!
factorádico:	2	2	0	0
	$2 \times 3!$	$+ 2 \times 2!$	$+ 0 \times 1!$	$+ 0 \times 0! = 16$

La representación decimal única de la permutación $\pi = (3, 4, 1, 2)$ es **16**.

1.3. Método de Kraskov, Stögbauer y Grassberger (KSG)

Un enfoque distinto a las estimaciones de las distribuciones de probabilidad, ya sea con histogramas, Kernel density estimation o ASO, es estimar la información mutua basada en la búsqueda de k primeros vecinos [20].

Supongamos dos espacios X, Y con cierta métrica, y sea Z el espacio $Z = (X, Y)$. Podemos ordenar para cada punto $z_i = (x_i, y_i)$ sus vecinos más cercanos por su distancia: $d_{i,j} = \|(z_i - z_j)\| : d_{i,j_1} \leq d_{i,j_2} \leq \dots$ Para ordenar el espacio Z se usa la norma máxima

$$\|z - z'\| = \max\{\|x - x'\|, \|y - y'\|\}, \quad (1.16)$$

mientras que para ordenar los subespacios X e Y se puede usar cualquier norma. Denotemos por $\epsilon(i)/2$ la distancia desde el punto z_i a su k -ésimo vecino, y $\epsilon_x(i)/2$ y $\epsilon_y(i)/2$ las distancias al mismo punto proyectadas en los subespacios X e Y , respectivamente. Entonces, $\epsilon(i) = \max\{\epsilon_x(i), \epsilon_y(i)\}$.

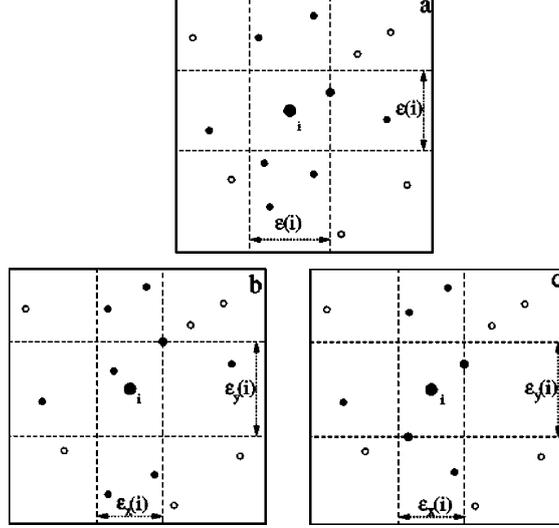


Figura 1.2: Panel(a): Determinación de $\epsilon(i)$, $n_x(i)$ y $n_y(i)$ en el primer algoritmo para $k = 1$ y un punto i fijo. En este ejemplo, $n_x(i) = 5$ y $n_y(i) = 3$. Paneles (b) y (c): Determinación de $\epsilon_x(i)$, $\epsilon_y(i)$, $n_x(i)$ y $n_y(i)$ en el segundo algoritmo para $k = 2$. El panel (b) muestra el caso en que $\epsilon_x(i)$ y $\epsilon_y(i)$ son determinados por el mismo punto, mientras que el panel (c) muestra el caso en que son determinados por distintos puntos.

En el primer algoritmo se cuenta el número $n_x(i)$ de puntos x_j cuya distancia al punto x_i es estrictamente menor que $\epsilon(i)/2$, y de forma similar para y en vez de x . En la figura 1.2 (a) se muestra este método. Como $\epsilon(i)$ es una variable aleatoria, y por consiguiente también lo son $n_x(i)$ y $n_y(i)$, denotamos por $\langle \dots \rangle$ al promedio sobre todos los puntos $i \in [1, \dots, N]$ sobre todas las realizaciones de las muestras aleatorias:

$$\langle \dots \rangle N^{-1} \sum_{i=1}^N E[\dots(i)] \quad (1.17)$$

La estimación de la información mutua es entonces

$$\mathbf{I}^{(1)}(X, Y) = \Psi(k) - \langle \Psi(n_x + 1) + \Psi(n_y + 1) \rangle + \Psi(N) \quad (1.18)$$

donde $\Psi(x) = \frac{d}{dx} \log(\Gamma(x))$ es la función digamma, o la derivada logarítmica de la función gamma. En el segundo algoritmo se reemplaza $n_x(i)$ y $n_y(i)$ por el número de puntos con $\|x_i - x_j\| \leq \epsilon_x(i)/2$ (figuras 1.2 (b) y (c)). La estimación de la información mutua es

$$\mathbf{I}^2(X, Y) = \Psi(k) - 1/k - \langle \Psi(n_x) + \Psi(n_y) \rangle + \Psi(N) \quad (1.19)$$

El algoritmo puede ser generalizado a más dimensiones utilizando la identidad 1.9. Aplicando la misma lógica que en dos dimensiones y encontrando los k primeros vecinos en la distribución conjunta $p(x, y, z)$, se obtiene

$$\mathbf{I}^{(1)}(X, Y|Z) = \Psi(k) - \langle \Psi(n_{xz}) - \Psi(n_{yz}) + \Psi(n_z) \rangle \quad (1.20)$$

donde se utilizó la norma máxima ϵ para encontrar los k primeros vecinos en una bola alrededor de (x, y, z) para cada punto, n_z es el número de primeros cuya distancia a (x, y, z) en el subespacio z es estrictamente menor a ϵ , y n_{xz} y n_{yz} son la cantidad de vecinos cuya distancia (en la norma máxima) a (x, y, z) es menor a ϵ en los subespacios (x, z) y (y, z) , respectivamente.

Mientras que para el algoritmo 2 se tiene que

$$\mathbf{I}^{(2)}(X, Y|Z) = \Psi(k) - \frac{2}{k} + \langle \Psi(n_z) - \Psi(n_{xz}) + \frac{1}{n_{xz}} - \Psi(n_{yz}) + \frac{1}{n_{yz}} \rangle \quad (1.21)$$

Si N es el largo de la muestra, el tiempo de cómputo en estos crece como $\mathcal{O}kN \log N$ si se utilizan algoritmos tipo k - d tree en la búsqueda de primeros vecinos, en vez de $\mathcal{O}(kN^2)$ [20]. En el presente trabajo sólo se utilizará la estimación 1.18 para la información mutua, ya que el cálculo de la transferencia de entropía buscando primeros vecinos en tres dimensiones usando la ecuación 1.20 ó 1.21 es muy costoso computacionalmente para las series temporales largas que se que se requieren para estimar las medidas de información en los sistemas complejos que se analizarán.

Como su nombre lo indica, el método está basado en la búsqueda de primeros vecinos en un espacio n -dimensional. Gran cantidad del tiempo de cómputo del algoritmo es utilizado en la búsqueda de primeros vecinos. El código está implementado enteramente en Python 3.7, y hace uso de la librería Scipy. El método que utiliza en la búsqueda es el k - d tree, una estructura espacial de particionamiento de datos para organizar puntos en un espacio k -dimensional. Si n es el largo de la distribución de puntos en un espacio bi-dimensional, el tiempo de cómputo del método es $\mathcal{O}(n \log(n))$, mientras que la búsqueda por fuerza bruta es $\mathcal{O}(n^2)$. El código fuente se encuentra en el proyecto [21].

1.4. Estimación de información mutua mediante redes neuronales (MINE)

[22] presentaron una forma de estimar la información mutua entrenando redes neuronales ¹ vía Basckpropagation. El resultado clave que utilizan los autores es la *representación dual* de la divergencia de Kullback-Leibler [24]:

Teorema (Representación de Donsker-Varadan): *La divergencia de Kullback-Leibler admite la siguiente representación dual:*

$$D_{KL}(P, Q) = \sup_{T: \Omega \rightarrow \mathbb{R}} [\langle T \rangle_P - \log(\langle e^T \rangle_Q)] \quad (1.22)$$

donde el supremo es tomado sobre todas las funciones T tal que los valores de expectativa son finitos.

Demostración: ver anexo de [22].

¹ver [23] para más información sobre redes neuronales.

Una consecuencia directa del Teorema es la siguiente cota para la estimación ([25]):

Lema: *Sea \mathcal{F} cualquier clase de funciones $T : \Omega \rightarrow \mathbb{R}$ que satisfacen las condiciones de integrabilidad del Teorema. Entonces se tiene el siguiente límite inferior:*

$$D_{KL}(P, Q) \geq \sup_{T \in \mathcal{F}} \langle T \rangle_P - \log(\langle e^T \rangle_Q) \quad (1.23)$$

La idea básica del *Estimador Neuronal de Información Mutua* (MINE, por sus siglas en inglés) es, usando la representación dual de la divergencia de Kullback-Leibler, elegir a \mathcal{F} como la familia de funciones $T_\theta : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ parametrizadas por una red neuronal de parámetros $\theta \in \Theta$. Se explota el límite

$$\mathbf{I}(X, Y) \geq \mathbf{I}_\Theta(X, Y) \quad (1.24)$$

donde \mathbf{I}_Θ es la *medida de información neuronal*, definida como

$$\mathbf{I}_\Theta(X, Y) = \sup_{\theta \in \Theta} [\langle T_\theta \rangle_{P_{X,Y}} - \log(\langle e^{T_\theta} \rangle_{P_X \otimes P_Y})] \quad (1.25)$$

Los valores de expectación son calculados usando muestras aleatorias de la probabilidad conjunta $P_{X,Y}$ y las marginales P_X, P_Y . Esta función puede ser maximizada mediante ascenso por gradiente.

Si P es una dada distribución de probabilidad, se denota por $P^{(n)}$ a la distribución de probabilidad empírica asociada a n muestras independientes e idénticamente distribuidas.

Definición: Sea $\mathcal{F} = \{T_\theta\}_{\theta \in \Theta}$ el conjunto de funciones parametrizadas por una red neuronal. MINE está definido como

$$\widehat{\mathbf{I}}(X, Y)_n = \sup_{\theta \in \Theta} [\langle T_\theta \rangle_{P_{XY}^n} - \log(\langle e^{T_\theta} \rangle_{P_X^{(n)} \otimes P_Y^{(n)}})] \quad (1.26)$$

Este método es fuertemente consistente. (ver [22] para la demostración de la consistencia del método).

La implementación de MINE es propuesta en el siguiente algoritmo:

Algoritmo 3 MINE

repeatExtraer b lotes de la distribución conjunta:

$$(x^{(1)}, y^{(1)}), \dots, (x^{(b)}, y^{(b)})$$

Extraer b muestras de la distribución marginal Y :

$$\bar{y}^{(1)}, \dots, \bar{y}^{(b)}$$

Evaluar el límite inferior:

$$\nu(\theta) \rightarrow \frac{1}{b} \sum_{i=1}^b T_{\theta}(x^{(i)}, y_{(i)}) - \log\left(\frac{1}{b} \sum_{i=1}^b e^{T_{\theta}(x^{(i)}, \bar{y}^{(i)})}\right)$$

Evaluar bias corrected gradients

$$\hat{G}(\theta) \rightarrow \tilde{\nabla}_{\theta} \nu(\theta)$$

Actualizar los parámetros estadísticos de la red:

$$\theta \rightarrow \theta + \hat{G}(\theta)$$

until convergencia

El código utilizado para estimar la MI con el método MINE está implementado en python3, utilizando la librería de aprendizaje automático *tensorflow* 1.14, y puede ser encontrado en el proyecto [26]. Si $\{x_t\}$ y $\{y_t\}$ son las series temporales de largo N con las cuales se quiere calcular la MI, la entrada de la red serán dos series $\{x_t\}_{in} = (\{x_t\}, \{x_t\})$, $\{y_t\}_{in} = (\{y_t\}, \text{shuffle}(\{y_t\}))$ de largo $2N$, donde "shuffle" significa mezclar aleatoriamente la serie y (esto se hace para estimar la distribución de probabilidad marginal a partir de la conjunta [22]). La primer capa consiste en dos subcapas lineales: capa x y capa y , totalmente conectadas a las entradas x e y respectivamente. La capa 2 está conectada a la capa x y a la capa y neurona a neurona, y su función de activación es un rectificador lineal (ReLU): $\text{ReLU}(x) = \max(0, x)$. O sea que: capa 2 = $\text{ReLU}(\text{capa } x + \text{capa } y)$. Las capas x , y y 2 tienen 20 neuronas cada una. La salida es una sola neurona totalmente conectada a la capa 2. En la figura 1.3 se muestra un diagrama de la red, donde las capas x , y y 2 tienen sólo dos neuronas (a modo de ilustración).

Aprovechando el resultado (1.23), se define la función de costo de la red como

$$E = - \left[\langle \text{output}(1 : 2N) \rangle - \log \left(\langle e^{\text{output}(2N:4N)} \rangle \right) \right] \quad (1.27)$$

o sea, $E = -\widehat{I}(x, y)_n$. La red se entrena en cada época con el optimizador Adam [27] con un ritmo de aprendizaje de 0,01.

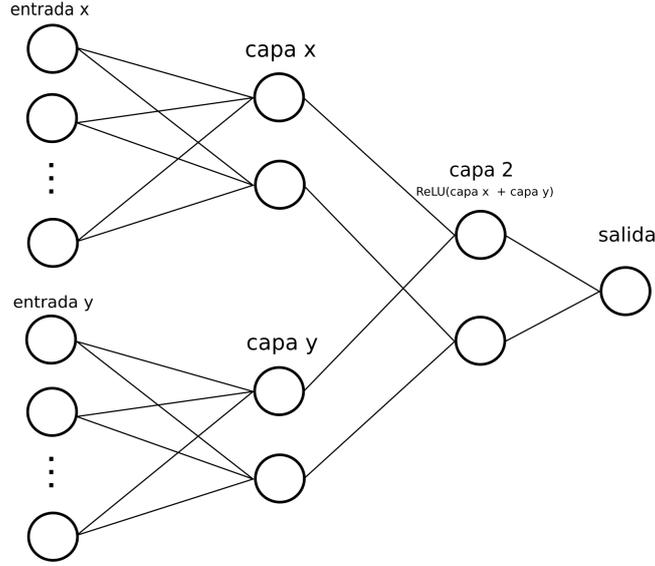


Figura 1.3: Estructura de la red neuronal utilizada en el algoritmo MINE.

Para evaluar la performance del MINE en la estimación de la información mutua, se analizará la convergencia del método utilizando series gaussianas, ya que se puede calcular la información mutua de forma analítica. Si se tienen n procesos gaussianos con una dada matriz de covarianza Σ y varianzas $\{\sigma_1, \sigma_2 \dots \sigma_n\}$, la información mutua entre todas las variables es

$$\mathbf{I}(X_1, X_2, \dots X_n) = -\frac{1}{2} \log \frac{|\Sigma|}{\sigma_1^2 \sigma_2^2 \dots \sigma_n^2} \quad (1.28)$$

donde $|\Sigma|$ denota el determinante de Σ .

En el caso especial dos de variables, la expresión 1.28 se simplifica a

$$\mathbf{I}(X_1, X_2) = -\frac{1}{2} \log(1 - \rho^2) \quad (1.29)$$

donde ρ es el coeficiente de correlación de Pearson entre X_1 y X_2 . Mientras más grande es la correlación entre las series, más grande es la información mutua.

Se generaron series gaussianas aleatorias de 5000 puntos (figura 1.4 a), con matriz de covarianza $\Sigma_{xx} = 1, \Sigma_{xy} = \Sigma_{yx} = 0,5$ y $\Sigma_{yy} = 1$ y se entrenó una red neuronal durante 200 épocas de entrenamiento, con realizaiones independientes de las series. El resultado se observa en la figura (1.4 b).

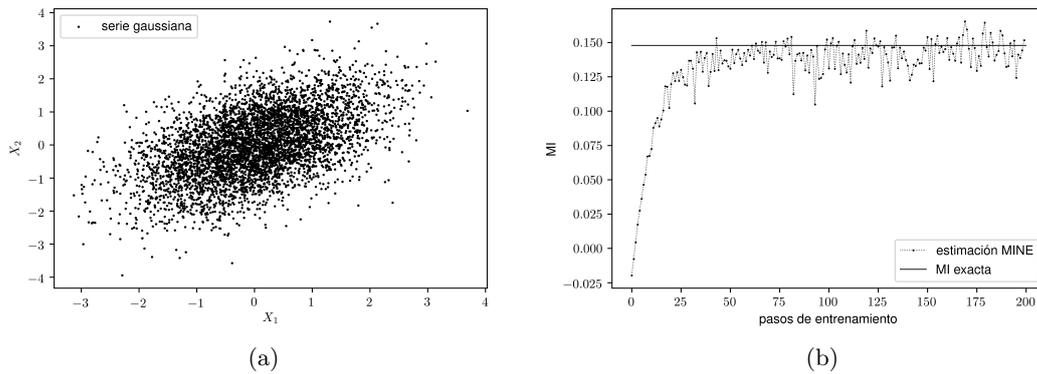


Figura 1.4: (a): series aleatorias gaussianas. (b): Información mutua calculada con el método MINE en cada época de entrenamiento, e información mutua exacta.

Para estimar el error de la medida se promediaron los valores de la MI desde las épocas 50 hasta 200 y se calculó la desviación estándar. El resultado se observa en la figura 1.5. El valor medio de la MI estimada (I_{mine}) no discrepa del valor exacto, pues cae dentro del intervalo $[I_{mine} - \Delta I_{mine}, I_{mine} + \Delta I_{mine}]$, donde ΔI_{mine} es la desviación estándar.

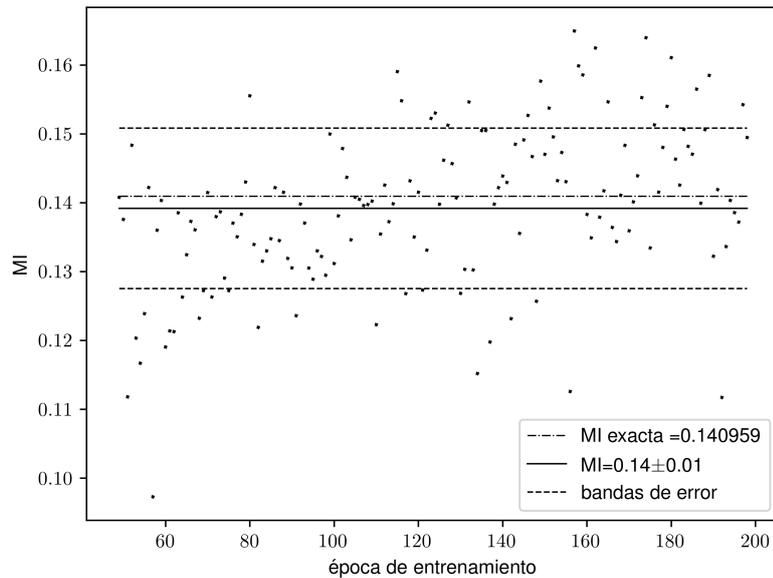


Figura 1.5: Información mutua de un proceso gaussiano calculada con MINE cuando el método converge, comparado con el valor de MI exacta.

2 Sistemas dinámicos

Potencialmente el objetivo de este trabajo sería aplicar estos cuantificadores de información directamente a variables de la atmósfera, en este sentido idealmente se podrían utilizar los modelos numéricos utilizados para el pronóstico meteorológico que se utilizan en los distintos centros de pronósticos. Sin embargo, para estudiar el comportamiento en un sistema más manejable y en el cual conozcamos exactamente su evolución y las interacciones entre las variables, se utilizarán sistemas dinámicos más sencillos pero que tienen ciertas características claves de la dinámica atmosférica. En este trabajo se usarán tres modelos:

- Lorenz 96: modela la evolución de una variable en un círculo de latitud.
- Modelo GOY: es un sistema de ecuaciones diferenciales ordinarias que emulan la turbulencia en 3d.
- Ecuación barotrópica de la vorticidad: Un conjunto de ecuaciones en derivadas parciales que es una forma límite de las ecuaciones de Navier-Stokes.

2.1. Modelo Lorenz-96

El primer sistema que será estudiado es el *Lorenz 96* [15]. Es un sistema dinámico no lineal, dado por un conjunto de N ecuaciones de las variables X_n :

$$\frac{dX_n}{dt} + X_{n-1}(X_{n-2} - X_{n+1}) + X_n = F \quad (2.1)$$

donde $n = 1, \dots, n$. Este conjunto de ecuaciones se representan en un dominio periódico, es decir $X_0 = X_N$, $X_{-1} = X_{N-1}$, $X_{-2} = X_{N-2}$.

Lorenz sugirió este modelo como un simple modelo atmosférico unidimensional en un círculo de latitud, usado para demostrar problemas de predicabilidad atmosférica. La física de la atmósfera está presente en este modelo sólo en el hecho de que hay un forzado externo F (la energía proveniente del sol), una disipación interna (viscosidad newtoniana), y los términos cuadráticos en conjunto conservan la energía del sistema $(X_1^2 + \dots + X_N^2)/2$. Se asume $N > 3$, y las variables se reescalan de tal manera que los coeficientes de los términos lineales y cuadráticos son iguales a la unidad, y se asume que este reescalado hace que la unidad temporal del modelo sea de 5 días.

Este sistema puede ser generalizado al *Lorenz 96 de dos escalas* [15]. El modelo consiste en dos sistemas del tipo 2.1 acoplados: uno con variables espaciales de gran amplitud y grandes longitudes de onda (lo que se denomina la "gran escala"), y otro sistema de más

variables de pequeña amplitud y longitudes de onda cortas (pequeña escala), simulando, por ejemplo, ondas de Rossby (gran escala) y ondas de gravedad (pequeña escala).

El modelo está definido por las ecuaciones

$$\frac{dX_n}{dt} + X_{n-1}(X_{n-2} - X_{n+1}) + X_n = F - \frac{hc}{b} \sum_{j=M/N(n-1)+1}^{nM/N} Y_j \quad (2.2)$$

$$\frac{dY_m}{dt} + cbY_{m+1}(Y_{m+2} - Y_{m-1}) + cY_m = \frac{hc}{b} X_{int[(m-1)/M/N]+1}$$

donde $n = 1, \dots, N$ y $m = 1, \dots, M$. Los dos conjuntos de ecuaciones se representan en un dominio periódico, es decir $X_0 = X_N$, $X_{-1} = X_{N-1}$, $X_{-2} = X_{N-2}$ y $Y_0 = Y_M$, $Y_1 = Y_{M+1}$, $Y_2 = Y_{M+2}$. Las ecuaciones (2.2) son básicamente las mismas pero en diferentes escalas. Tienen una constante de acople h entre las escalas, coeficiente de proporción espacial y temporal b y c respectivamente. Las variables de la pequeña escala son forzadas por la variable local (más cercana) de la gran escala, y las variables de la gran escala son forzadas por el forzado externo F y por todas las variables de la pequeña escala que están alrededor de la variable de la gran escala considerada. La imagen (2.1) muestra una representación gráfica del modelo.

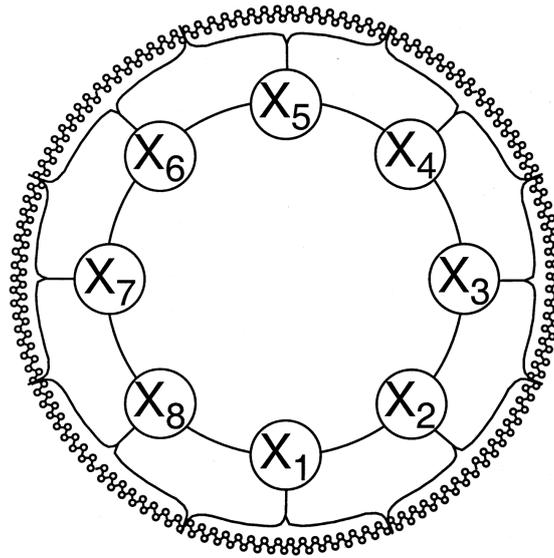


Figura 2.1: Representación gráfica del *Lorenz 96 de dos escalas*, con $N = 8$ variables de la gran escala y $M = 256$ variables de la pequeña escala (32 variables de la pequeña escala por cada una de las variables de la gran escala).

Para resolver numéricamente el sistema dinámico de una escala se utilizó el método de Runge-Kutta de cuarto orden, con paso de integración $h = 0,001$, tiempo de observación $t_{obs} = 0,05$, número de variables $N = 40$ y forzado externo $F = 8$. Al modelo de dos escalas se lo resolvió con el mismo tiempo de resolución y observación que el de una

escala, y se seleccionaron los parámetros como $b = c = 1$, $h = 10$, $N = 8$, $M = 256$ (32 variables de la pequeña escala acopladas a una de la gran escala) y $F = 16$.

El comportamiento de las series temporales de estos sistemas dinámicos depende mucho del forzado externo F del modelo. En el caso del Lorenz 96 de una escala, las soluciones pueden ser caóticas para $F = 8$ (figura 2.2 abajo), periódicas para $F = 3$ (figura 2.2 medio), y disipativas para $F < 1$ (figura 2.2 arriba).

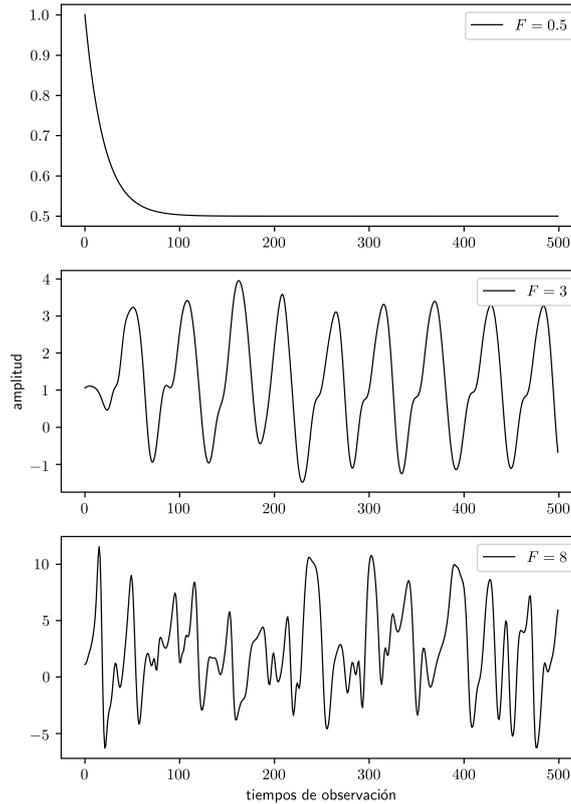


Figura 2.2: Varios comportamientos del Lorenz 96. Arriba: comportamiento disipativo, $F = 0,5$. Centro: comportamiento periódico, $F = 3$. Abajo: comportamiento caótico, $F = 8$.

En el caso del Lorenz 96 de dos escalas, el comportamiento de la pequeña escala es predominantemente disipativo, por lo que se requieren forzados aún mayores que en el modelo de una escala para asegurar caos en todo el sistema. En la figura 2.3 se muestran series temporales del sistema para varios forzados externos. Para $F = 16$ las solución del modelo es caótica (figura 2.3, abajo), mientras que para $F = 8$ la solución es periódica (figura 2.3, centro), y finalmente disipativas para $F = 0,5$ (2.3, arriba). Una descripción

más detallada de las soluciones de los modelos de Lorenz-96 y su impacto en las medidas del análisis simbólico ordinal se encuentran en [14].

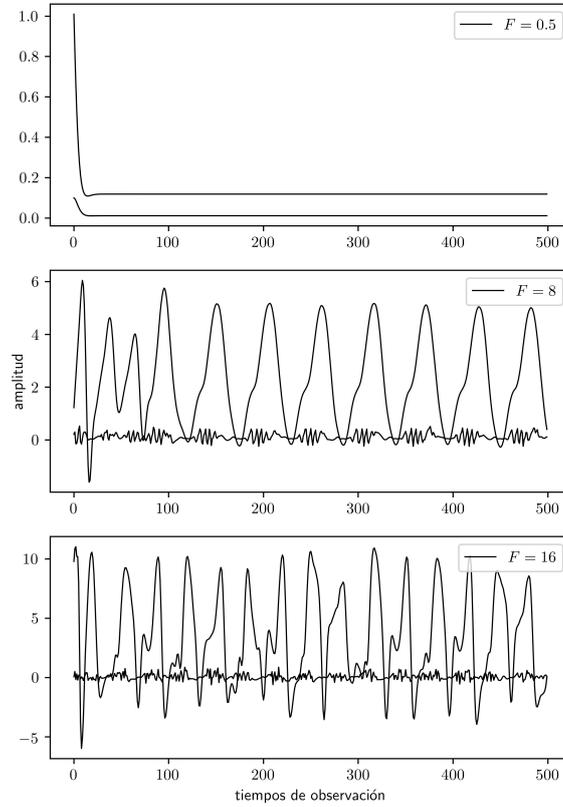


Figura 2.3: Varios comportamientos del Lorenz 96 de dos escalas. Arriba: comportamiento disipativo, $F = 0,5$. Centro: comportamiento periódico, $F = 8$. Abajo: comportamiento caótico, $F = 16$.

Teniendo en cuenta los comportamientos de los sistemas, se optó por realizar los experimentos con series temporales del modelo sólo en el régimen caótico, con $F = 8$ en el caso del Lorenz 96 de una escala, y $F = 16$ en el caso del modelo de dos escalas. Cabe aclarar que para trabajar con medidas de información es esencial que el sistema sea caótico [28]. En el capítulo 3 se aplicarán las medidas de información mutua I_{X_i, X_j} y transferencia de entropía $T_{X_j \rightarrow X_i}$ para detectar interacciones entre variables del sistema y el rango temporal donde éstas ocurren. Para eso se calculará la MI entre una variable X_i a tiempo presente y otra X_j a tiempo pasado. Esto se logra desfasando las series temporales en un tiempo τ , que estará medido en tiempos de observación. Con esta misma metodología se calculará la transferencia de entropía entre variables en función

del lag temporal de las series.

2.2. Modelo de Gledzer-Ohkitana-Yamada (GOY)

Desde la teoría de los sistemas dinámicos caóticos, la turbulencia tridimensional completamente desarrollada en las ecuaciones de Navier-Stokes puede ser interpretada como la manifestación de un atractor extraño de muy alta dimensionalidad en un espacio de fases de dimensión infinita. Dada la alta dimensionalidad del problema ($\mathcal{O}(10^6)$), no es posible utilizar directamente la metodología de la teoría del caos.

[12] proponen un modelo (denominado GOY, de ahora en adelante) de turbulencia totalmente desarrollada, el cual satisface la Ley de Kolmogorov $k^{-5/3}$ del espectro de energía en el rango inercial (Ver anexo para más información). La variable que evoluciona en el tiempo es la velocidad $V_h(t)$ en el h -ésimo sitio de una grilla unidimensional. Cada punto de la grilla evoluciona como $V_h = V_h(t)$. Con respecto a la dependencia espacial de la variable, se puede realizar una transformada de Fourier: de las variables reales $V_h(t)$, se construye un conjunto de ecuaciones $u_n = u_n(t)$, donde u_n es la velocidad del n -ésimo modo de Fourier del campo de la velocidad, caracterizado por un número de onda k_n . Este n -ésimo número de onda está dado por la relación

$$k_n = k_0 q^n \quad (2.3)$$

donde k_0 es el modo más bajo (fundamental) y q es un coeficiente de magnificación que relaciona al n -ésimo con el $(n+1)$ -ésimo número de onda mediante la relación $k_{n+1} = qk_n$. La elección de los autores es $q = 2$, lo que significa que cuando las cascadas ocurren, el tamaño de los remolinos se reduce a la mitad desde u_n a u_{n+1} .

Cada modo $u_n(t)$ es un proceso físico, y todos los procesos interaccionan entre sí. El campo de velocidades en el espacio físico está supuestamente gobernado por las ecuaciones de Navier-Stokes, cuyas no linealidades llevan al acoplamiento entre los distintos modos. El sistema no está aislado, sino que un forzado externo f_n actúa sobre el modo $n = 4$ de la forma $f_n = \delta_{4,n} n(1+i)f$, donde i es la unidad imaginaria y f una constante de amplitud. El sistema de ecuaciones diferenciales ordinarias que gobierna la evolución de u_n es

$$\begin{aligned} \frac{du_n}{dt} &= i(k_n u_{n+1} u_{n+2} - \frac{1}{2} k_{n-1} u_{n+1} u_{n-1} - \frac{1}{2} k_{n-2} u_{n-1} u_{n-2})^* - \nu k_n^2 u_n + f_n, \\ f_n &= \delta_{4,n} (1+i)f \end{aligned} \quad (2.4)$$

donde $n = 1, 2, \dots$ y x^* denota el conjugado de x . Cada modo u_n está coplado con u_{n+1} , u_{n+2} y u_{n-1} de forma cuadrática, y con sigo mismo mediante un término disipativo $-\nu k_n^2 u_n$ que imita a la *hiperviscosidad*: una disipación que es sólo significativa en los números de onda altos, dejando a los números de onda bajos esencialmente conservativos.

El modelo GOY está gobernado por un sistema de ecuaciones diferenciales ordinarias *rígidas* (*stiff equations*, en inglés), es decir, las soluciones del modelo varían rápidamente

con el tiempo, y los métodos numéricos usuales son inestables, a menos que se tome un paso de resolución temporal muy pequeño. El código utilizado está disponible en [29]. Utiliza versiones modificadas del método de Runge-Kutta de cuarto orden para sistemas rígidos, basadas en el método de diferencias de tiempo exponencial [30].

Se seleccionaron los parámetros del modelo de acuerdo a [12]: $f = 5 \times 10^{-3}(1 + i)$ (amplitud del forzado en el modo 4) y $\nu = 10^{-7}$ (coeficiente de hiperviscosidad). El paso de integración es $h = 10^{-3}$ y el tiempo de observación de las series es $t_{\text{obs}} = 0,1$.

En la figura 2.4 se muestra la evolución de los modos del modelo. Los modos bajos tienen más amplitud y período largo, mientras que los modos altos tienen una amplitud pequeña y período corto.

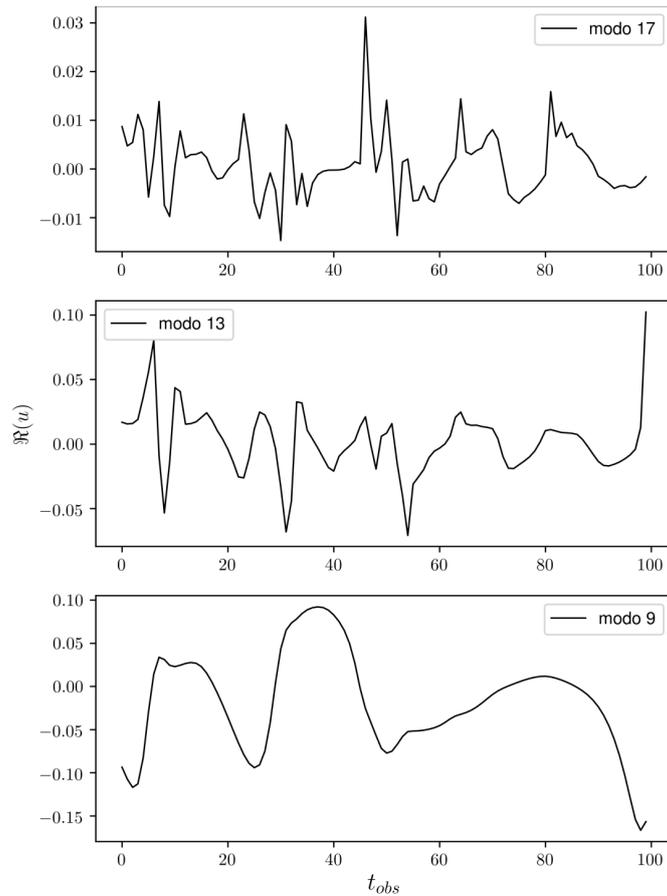


Figura 2.4: Series temporales de los modos 17, 13 y 9 de la velocidad en el modelo GOY.

Al ser un modelo sintético de turbulencia 3D, el espectro de la energía sigue la ley de Kolmogorov $E(k) \propto -5/3$. La figura 2.5 muestra el espectro promedio en una serie de 50,000 puntos, comparado con la ley de potencias $k^{-5/3}$ en el rango inercial (donde la hiperviscosidad es despreciable).

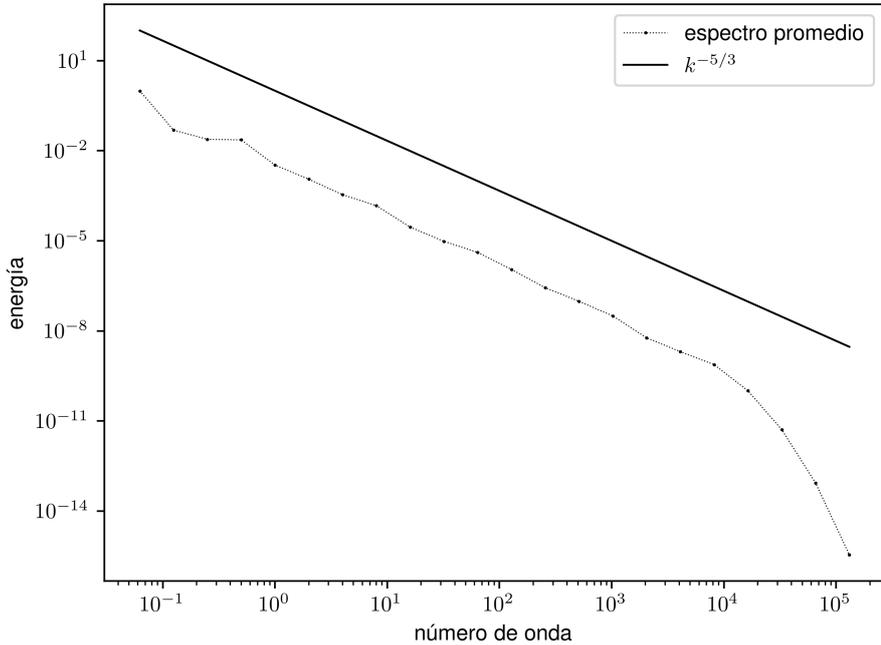


Figura 2.5: Espectro promedio de la energía para series temporales de 50,000 puntos.

2.3. Ecuación barotrópica de la vorticidad

Una peculiaridad que tiene la turbulencia de gran escala en la atmósfera es que es esencialmente bidimensional. La existencia de la fuerza de Coriolis, estratificación y el pequeño ancho del medio geofísico restringe severamente la velocidad vertical en flujos de gran escala, los cuales tienden a ser cuasigeostroficados, con la fuerza de Coriolis balanceando el gradiente de presión horizontal. Dado que el estiramiento de vórtices, el mecanismo clave la turbulencia tridimensional ordinaria transfiere energía desde las grandes escalas hacia las escalas pequeñas, no existe en dos dimensiones, la dinámica de turbulencia geostrofica es muy diferente a la turbulencia tridimensional de pequeña escala.

Un resultado clave [31] muestra que la energía va a escalas grandes, mientras que la enstrofia va hacia las escalas pequeñas. Si la energía es inyectada al sistema con una velocidad ϵ , entonces el espectro de energía en la región de cascada de energía es $S(k) \propto \epsilon^{2/3} k^{-5/3}$. Un argumento dimensional muestra que el espectro de energía en la región de cascada de enstrofia es de la forma $S(k) \propto \alpha^{2/3} K^{-3}$, donde α es el flujo de enstrofia hacia los números de onda altos. Se puede demostrar que el espectro de la enstrofia es $Z(k) = k^2 S(k)$ [32]. En la figura 2.6 se muestran las dos regiones inerciales donde ocurren las cascadas.

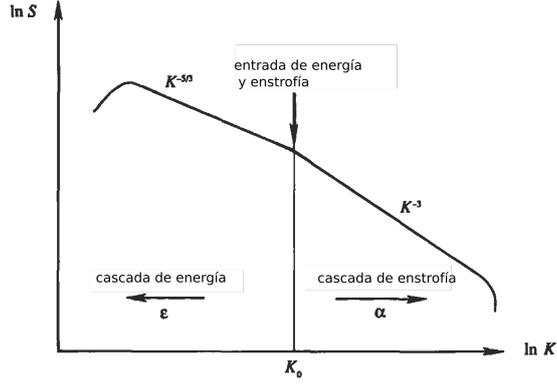


Figura 2.6: Cascadas de energía y enstrofia en la turbulencia bidimensional.

Un conjunto de ecuaciones suficientemente generales pueden modelar lo anteriormente dicho es el siguiente:

$$\frac{\partial}{\partial t}(\zeta - \gamma^2 \psi) + \mathbf{J}(\psi, \zeta - \gamma^2 \psi + \beta y) = f - \nu_0 \zeta + \nu_2 \nabla^2 \psi - \nu_4 \nabla^4 \zeta \quad (2.5)$$

$$\zeta = \nabla^2 \psi, \quad u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x} \quad (2.6)$$

donde x , y y t son las cordenadas Este, Norte y temporal, ψ es la función corriente, u y v son las velocidades en las direcciones Este y Norte y ζ es la vorticidad. γ es la inversa del radio de curvatura de la superficie, β es el gradiente del parámetro de Coriolis, f es una función de forzado y los términos ν_i son coeficientes de fricción, correspondientes a la fricción de Rayleigh, viscosidad Newtoniana e hiperviscosidad, correspondientes a $i = 0, 2, 4$, respectivamente. J es el operador Jacobiano en x e y .

Cuando el forzado y todos los términos difusivos son cero, el sistema conserva la energía y la enstrofia potencial:

$$E \equiv \frac{1}{2} \frac{1}{(2\pi)^2} \int \int dx dy (u^2 + v^2) \quad (2.7)$$

$$Z \equiv \frac{1}{(2\pi)^2} \int \int dx dy (\zeta - \gamma^2 \psi)^2$$

Un caso particular de las ecuaciones 2.6 es la *ecuación barotrópica de la vorticidad*:

$$\gamma = \beta = \nu_0 = \nu_2 = 0 \frac{\partial \zeta}{\partial t} + \mathbf{J}(\psi, \zeta) = f - \nu_4 \nabla^4 \zeta \quad (2.8)$$

El código utilizado para resolver la ecuación barotrópica de la vorticidad 2.8 puede ser encontrado en [33], el cual calcula la vorticidad en una grilla cuadrada. Para eso, utiliza métodos espectrales para expandir en series de fourier a la función corriente:

$$\Psi = \sum A_k e^{i(\vec{k} \cdot \vec{r})} \quad (2.9)$$

por lo que la velocidad se calcula de forma analítica a partir de 2.9:

$$\frac{\partial \Psi}{\partial x} = ik_x \Psi \quad \frac{\partial \Psi}{\partial y} = ik_y \Psi \quad (2.10)$$

Una vez calculada la velocidad en el espacio de Fourier, se computa el Jacobiano $\mathbf{J}(\Psi, \zeta)$, se calcula el forzado en cada tiempo de la forma $A_{i,j} = J'_{i,j} + F_{i,j}$, donde los índices i, j barren los valores de los números de onda en la grilla, y se aplica el método de Adams-Bashforth para hacer un paso temporal en la vorticidad. En cada paso de resolución se aplica a la vorticidad un forzado aleatorio uniforme en un cuadrado en el espacio de Fourier entre los números de onda $14 < k_x < 20$ y $14 < k_y < 20$ para que el modelo no sea enteramente disipativo.

El método utilizado para representar la viscosidad en los números de onda altos es el *filtro de Smith* [34]:

$$F_\nu = \begin{cases} \exp[-a(k - k_{cut})^s], & \text{si } k > k_{cut} \\ 1, & \text{si } k \leq k_{cut} \end{cases} \quad (2.11)$$

el cual aplica una disipación exponencial a los números de onda altos (a partir de $k_{cut} = k_{30}$). Se tomó el valor $s = 8$, y a tal que $F(k_{max}) = (1 + 4\pi/N)^{-1}$, lo que da

$$a = \frac{\ln(1 + 4\pi/N)}{(k_{max} - k_{cut})^s}$$

El modelo se inicializa con la *condición inicial de McWilliams* [35]:

$$\zeta \propto k \left[1 + \frac{k}{k_0} \right]^{-1} \quad (2.12)$$

donde $k > 0$ es el número de onda radial y k_0 es el máximo de la función de forzado. Se eligió $k_0 = 15$.

Para asegurar la estabilidad numérica, el código utiliza un método adaptativo tal que el número de Courant $C = \frac{u\Delta t}{\Delta x}$ se mantiene entre 0,4 y 0,6.

El tamaño de la grilla usada es 128×128 . Considerando la cantidad de variables del sistema, es muy difícil obtener información relevante con las medidas de información. Una forma de evitar ese problema es trabajar en el espacio de Fourier sólo con el número de onda radial. Así, en cada tiempo de observación ($t_{obs} = 0,25$), se calcula la energía y enstrofía 2.7 en el espacio de Fourier y se las integra alrededor del ángulo polar para obtenerlas en función del número de onda radial $E = E(k)$ y $Z = Z(k)$. De esta manera, el problema se reduce a considerar el acoplamiento entre N_k números de onda (ver [36] para ver el código fuente para la obtención de los espectros). Este acoplamiento es el que brinda información de las cascadas de energía y enstrofía en sistemas turbulentos. Cada serie temporal obtenida del modelo tiene un largo de 50000 puntos.

En la figura 2.7 se muestran los campos de vorticidad en el espacio físico para (a) 1, (b) 25, (c) 50 y (d) 100 tiempos de observación, propios de un flujo turbulento en dos dimensiones.

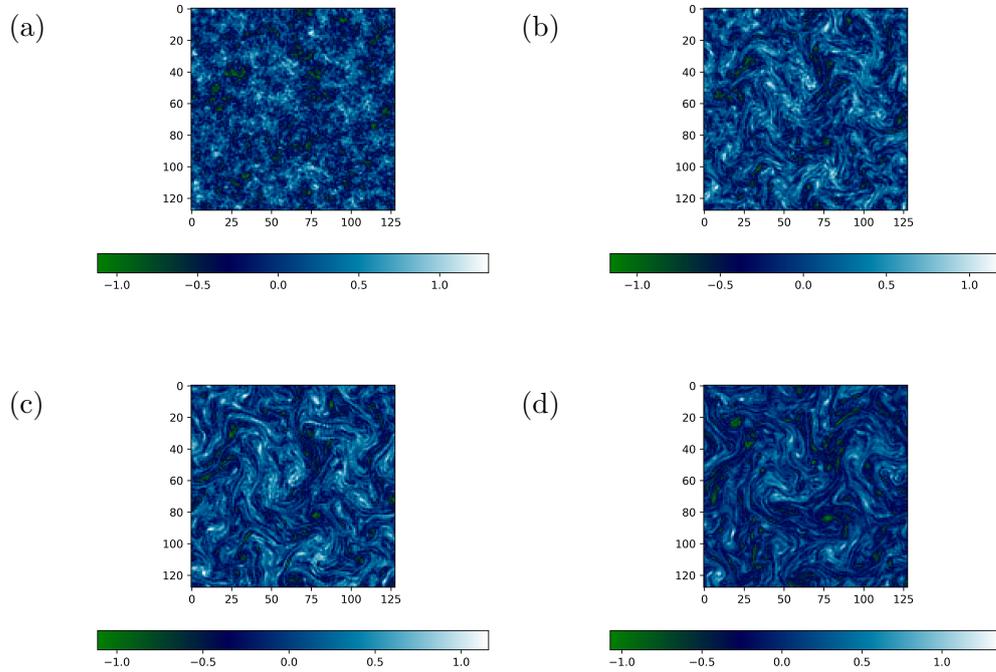


Figura 2.7: Campos de vorticidad en el espacio físico del modelo barotrópico para los tiempos (a) $t_{obs} = 1$, (b) $t_{obs} = 25$, (c) $t_{obs} = 50$ y $t_{obs} = 100$.

En la figura 2.8 se muestran los modos 5, 20 y 30 de la energía y la enstrofía en función del tiempo. Los modos con números de onda bajos tienen mayor amplitud y menor frecuencia y los modos con números de onda altos tienen pequeña amplitud y alta frecuencia.

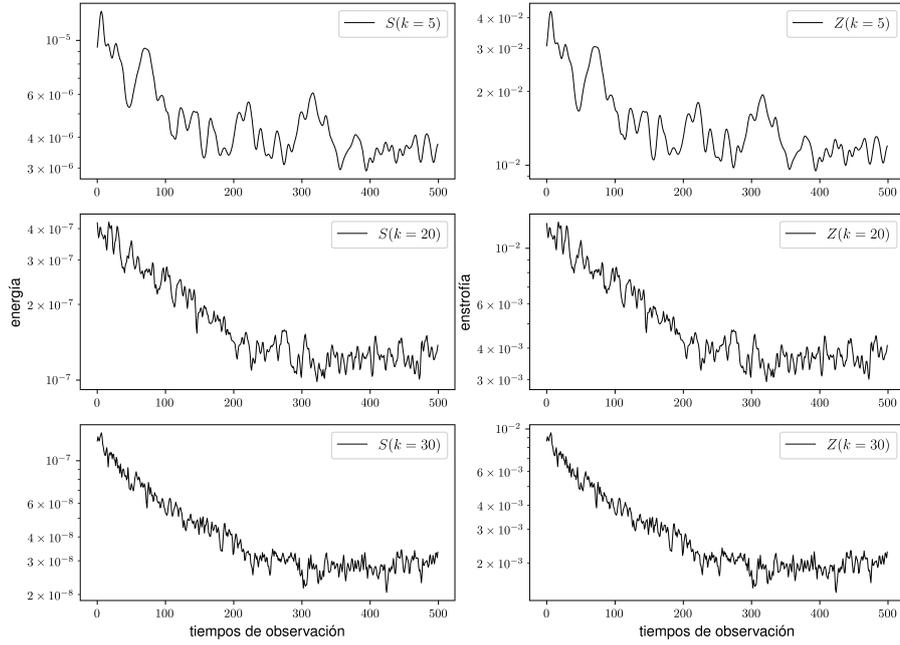


Figura 2.8: Izquierda: series temporales de los modos 5, 20 y 30 de la energía (izquierda) y enstrofia (derecha) en función del tiempo en el modelo barotrópico.

En la figura 2.9 a se muestra el espectro de la energía para varios tiempos de observación, comparado con el espectro teórico del modelo $S(k) \propto k^{-3}$, propio del rango inercial de la cascada directa de enstrofia (ver figura 2.6) de la turbulencia en dos dimensiones, mientras que en la 2.9 b se muestra el espectro de la enstrofia para varios tiempos de observación, comparado con el espectro teórico del modelo $Z(k) = k^2 S(k) \propto k^{-1}$.

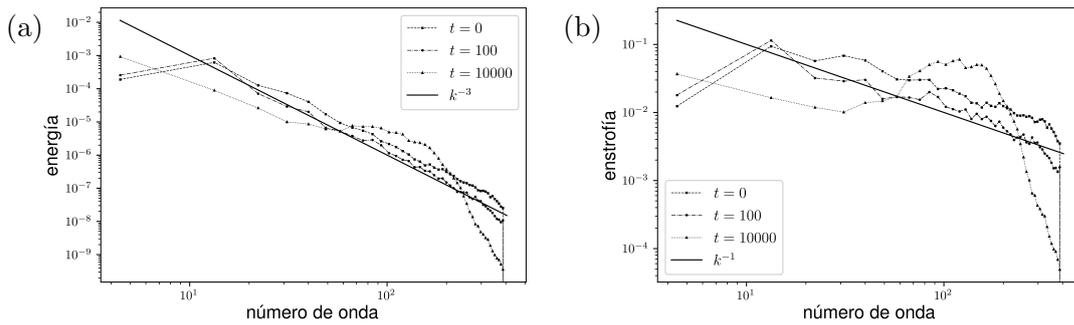


Figura 2.9: Espectro de (a) energía y (b) enstrofia en el modelo barotrópico para varios tiempos.

3 Experimentos con el Lorenz 96

3.1. Análisis simbólico ordinal

En esta sección se aplican las medidas de información mediante el uso del análisis simbólico ordinal a las series temporales de los modelos Lorenz-96 de una y dos escalas. Un primer análisis se enfoca en estudiar los acoplamientos espaciales entre las variables del sistema. En el caso del modelo de una escala, se calculó la información mutua entre la variable X_{20} a tiempo presente y el resto de las variables X_i , $i = 1, \dots, 40$ para los tiempos pasados $\tau = 0, 5$ y 10 . El largo de las series temporales es $n_{dat} = 50,000$. Los valores de información mutua obtenidos se muestran en la figura 3.1. Cada punto de las figuras 3.1a-c es el valor numérico de $\mathbf{I}(X_{20}^t, X_i^{t-\tau})$ en función de la variable i . A tiempo presente ($\tau = 0$) el máximo de MI está en $i = 20$, o sea, cuando se compara la variable X_{20} consigo misma (figura 3.1 a). A $\tau = 5$, la interacción con las variables a la izquierda de X_{20} se hace más significativa, aunque el máximo de MI sigue en X_{20} . A $\tau = 10$, es la variable X_{17} la que presenta más correlación con el presente de X_{20} , Aunque la correlación de X_{20} con su pasado sigue siendo significativa.

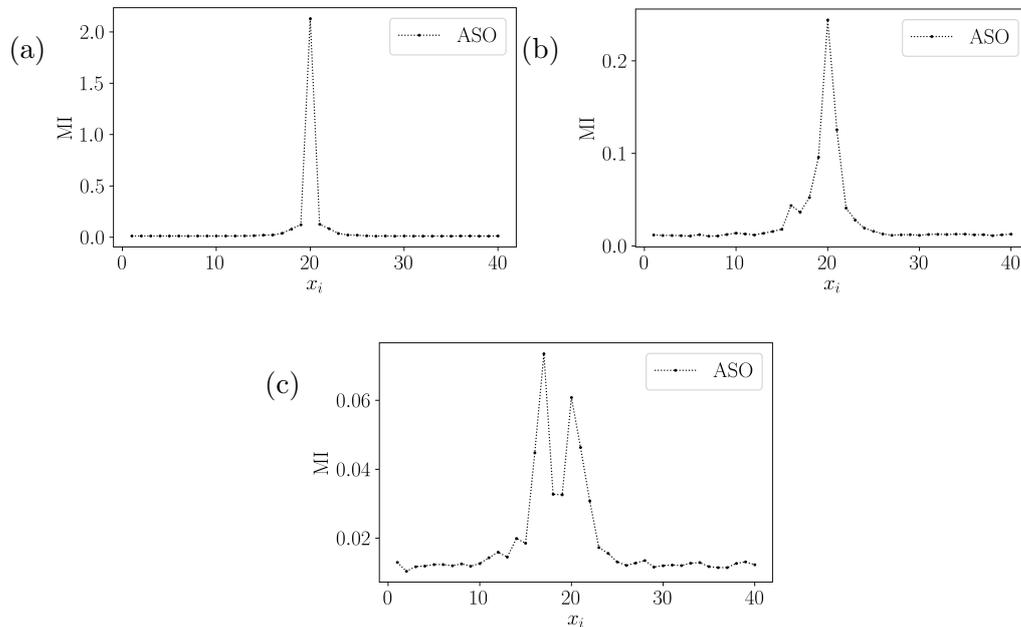


Figura 3.1: Información mutua con el método de BP de la variable X_{20} con todas las variables X_i del Lorenz 96 de una escala para (a) $\tau = 0$, (b) $\tau = 5$ y (c) $\tau = 10$.

Una vez determinada la direccionalidad de la información, el segundo análisis consiste en calcular la información mutua entre la variable X_{20} a tiempo presente con X_{17} , X_{18} , X_{20} y X_{21} a tiempo pasado en función del retardo τ de las series. Cada punto de la figura 3.2(a) es el valor numérico de $\mathbf{I}(X_{20}^t, X_i^{t-\tau})$, con $i = 17, 18, 20$ y 21 . Se observa que, cuando $\tau < 8$, la correlación de X_{20} con su propio pasado es siempre mayor que la correlación con el pasado de sus variables vecinas. Para los tiempos pasados $\tau < 7$, hay una alta correlación entre X_{20} con X_{21} , mientras que para $\tau = 13$ el pasado de X_{17} muestra mucha correlación con X_{20} . Finalmente, para $\tau = 19$ es la variable X_{18} que tiene más influencia sobre X_{19} .

También se calculó la TE desde X_{17} , X_{18} , X_{19} y X_{21} a tiempos pasados hacia el presente de X_{20} en función del lag temporal τ . En la figura 3.2(b) se muestran los valores numéricos de $\mathbf{T}_{X_i^{t-\tau} \rightarrow X_{20}^t}$, con $i = 17, 18, 19$ y 21 . La variable X_{19} es la que tiene más causalidad sobre X_{20} a $\tau = 5$, mientras que a $\tau = 6$ es X_{21} quien más influye sobre X_{20} . A $\tau = 8$ se observa la máxima interacción de X_{18} , la cual es más débil que los otros máximos de TE, mientras que para tiempos más lejanos ($\tau = 10 - 14$) se observa una interacción sostenida de X_{17} con X_{20} .

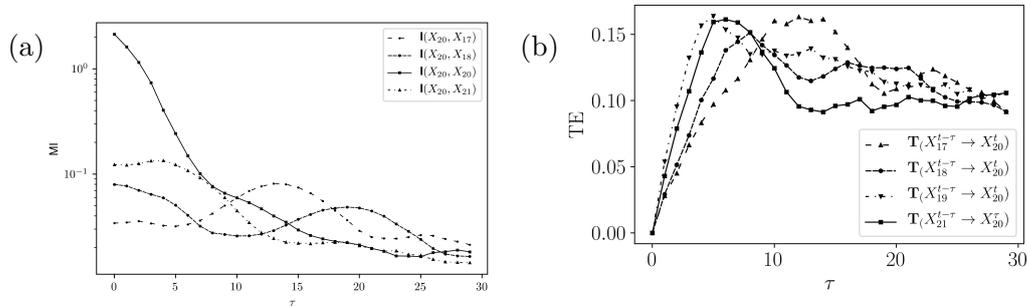


Figura 3.2: (a) Información mutua con el método de BP de la variable X_{20} con X_{17} , X_{18} , X_{20} y X_{21} en función de τ . (b) Transferencia de entropía desde los modos X_{17} , X_{18} , X_{19} y X_{21} hacia X_{20} en función de τ .

El error al estimar la MI y TE con el ASO se lo estimó de la siguiente forma: se calculó $\mathbf{I}(X_{20}^t, X_{19}^t)$ (MI entre X_{20} y X_{19} a tiempo presente) y $\mathbf{T}_{X_{19}^{t-\tau} \rightarrow X_{20}^t}$ (TE desde X_{19} a X_{20}) con $D = 5$ y $\tau = t_{obs} = 0,05$ para 300 realizaciones independientes de las series temporales y se calculó el promedio y la desviación estándar de las medidas. Los resultados se muestran en la figura 3.3. El error estadístico asociado al estimar la MI con el método de BP es de 0,3%, mientras que el de la TE es de 0,1%.

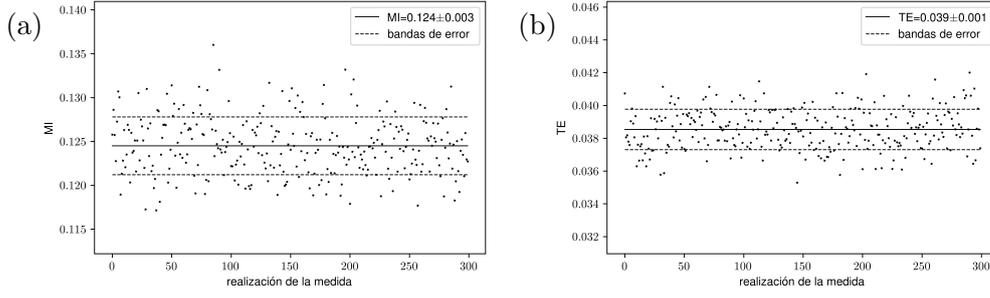


Figura 3.3: 300 realizaciones de (a) $\mathbf{I}(X_{20}^t, X_{19}^t)$ y (b) $\mathbf{T}_{X_{19}^{t-\tau} \rightarrow X_{20}^t}$ con el ASO.

Para determinar la utilidad de las medidas de información para encontrar acoplamientos entre las escalas de sistemas, se calculó la información mutua entre la variable X_4 de la gran escala y todas las Y_j , $j = 1, \dots, 256$ variables de la pequeña escala. El largo de las series temporales es $n_{dat} = 50,000$ puntos. El resultado se muestra en la figura 3.4.

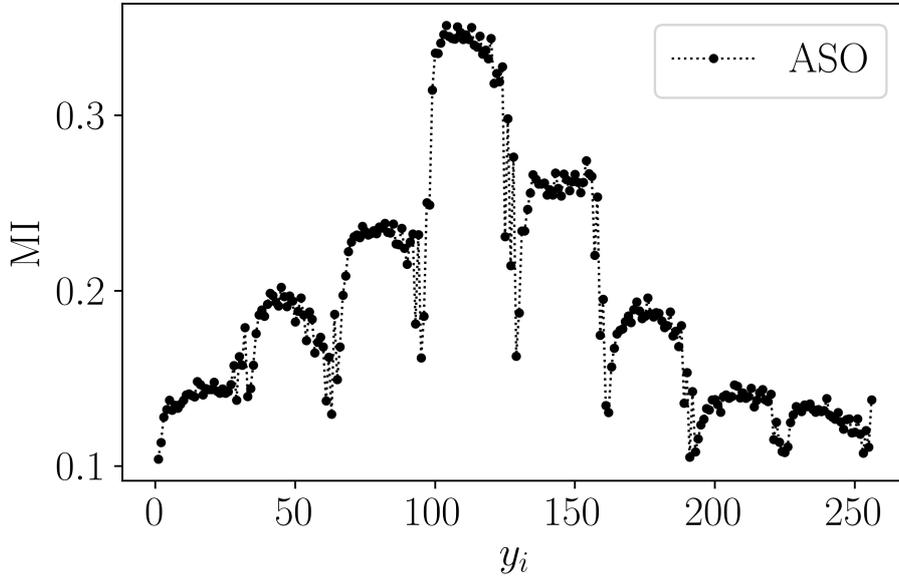


Figura 3.4: información mutua entre la variable X_4 de la gran escala y todas las 256 variables Y_j de la pequeña escala, calculada con el ASO.

Se observa claramente que la información mutua depende mucho de la variable de la pequeña escala involucrada (ver ecuación 2.2): las 32 variables Y_i , $i = 97, \dots, 128$, que sumadas actúan como forzado para X_4 , están más correlacionadas con X_4 debido a que estas variables de pequeña escala tienen un acoplamiento directo en las ecuaciones con la variable de gran escala considerada, mientras la información mutua desciende

sustancialmente en grupos de variables de pequeña escala que son forzadas por otras variables de la gran escala, cuyo acoplamiento es entonces indirecto.

El acoplamiento entre procesos de distintas escalas en la atmósfera es muy dificultoso de cuantificar. En principio, en este sistema prototipo los resultados muestran que las medidas de información basadas en el análisis simbólico ordinal pueden brindar información útil, bajo la hipótesis de que se posean series temporales suficientemente largas como para obtener distribuciones estacionarias.

3.2. Método de KSG

Para estudiar los acoplamientos espaciales entre las variables del sistema Lorenz-96 de una escala, se calculó la información mutua entre la variable X_{20} a tiempo presente y el resto de las variables X_i , $i = 1, \dots, 40$ para los tiempos pasados $\tau = 0, 5$ y 10 . El largo de las series temporales es $n_{dat} = 50,000$. No se utilizaron series temporales largas como en el análisis simbólico ordinal porque (a) el método es más demandante computacionalmente y (b) con esa cantidad de puntos se consiguen series temporales suficientemente estacionarias para los requerimientos del método KSG, logrando buenas estimaciones de MI. En la figura 3.5a-c se muestran los valores numéricos de $\mathbf{I}(X_{20}^t, X_i^{t-\tau})$ con $\tau = 0, 5$ y 10 respectivamente, en función de la variable i .

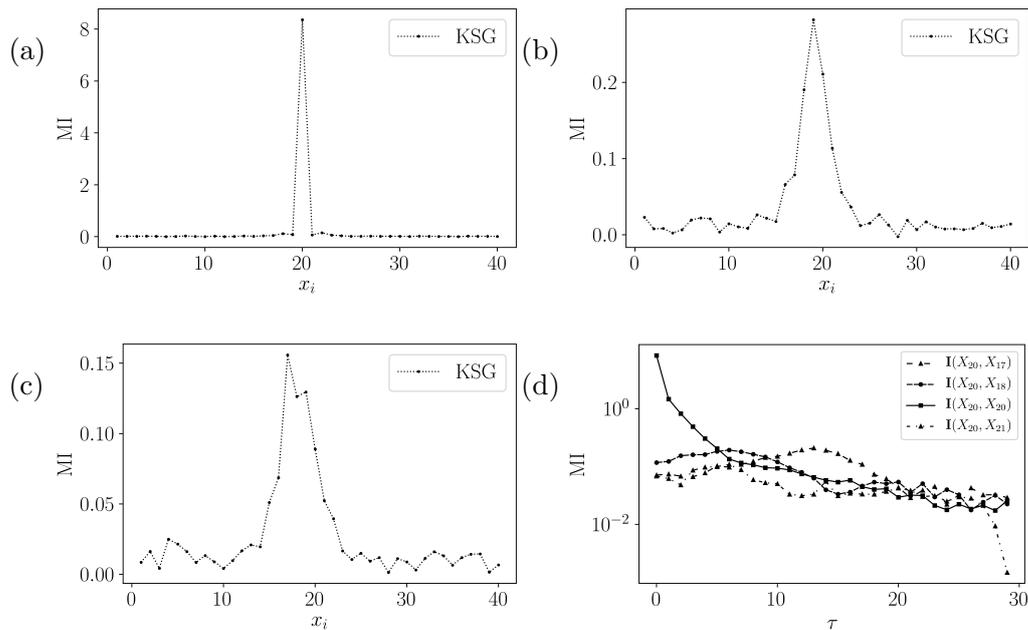


Figura 3.5: Información mutua con el método de KSG de la variable X_{20} con todas las variables X_i del Lorenz 96 de una escala para (a) $\tau = 0$, (b) $\tau = 5$ y (c) $\tau = 10$. En (d) se muestra la información mutua de la variable X_{20} con X_{17} , X_{18} , X_{20} y X_{21} en función de τ .

Cuando se calcula la MI sin desfases temporales ($\tau = 0$) entre las series, el máximo

de MI está en $i = 20$. A $\tau = 5$, la variable X_{19} a tiempo pasado tiene más correlación con X_{20} que el mismo pasado de X_{20} : el máximo de MI se corre una variable hacia la izquierda. A $\tau = 10$, la variable X_{17} está correlacionada con el presente de X_{20} mucho más que el pasado de la variable X_{20} , mostrando claramente la direccionalidad del flujo de información desde las variables de la izquierda hacia las de la derecha.

En la figura 3.5(d) se muestra la información mutua entre el presente de la variable X_{20} con el pasado de X_{17} , X_{18} , X_{20} y X_{21} en función del retardo τ de las series temporales. Se observa que la correlación de X_{20} con su pasado es siempre mayor que la correlación con sus variables vecinas para tiempos $\tau < 5$. Para tiempos pasados mayores, la información mutua con la variable X_{18} es máxima, mientras que en $\tau = 13$ es la variable X_{17} la que tiene mayor impacto sobre el presente de X_{20} . La correlación del pasado de X_{21} no es significativa comparada con las otras variables.

El error al estimar la MI con el método de KSG se lo estimó de la misma forma que el error con el método ASO: se calculó $\mathbf{I}(X_{20}^t, X_{19}^t)$ (MI entre X_{20} y X_{19} a tiempo presente) con $k = 3$ primeros vecinos y largo de las series temporales de 15,000 puntos, para 300 realizaciones independientes de las series temporales, y se calculó el promedio y la desviación estándar de las medidas. Los resultados se muestran en la figura 3.6. El error estadístico asociado al estimar la MI con el método de KSG es de 0,9 %.

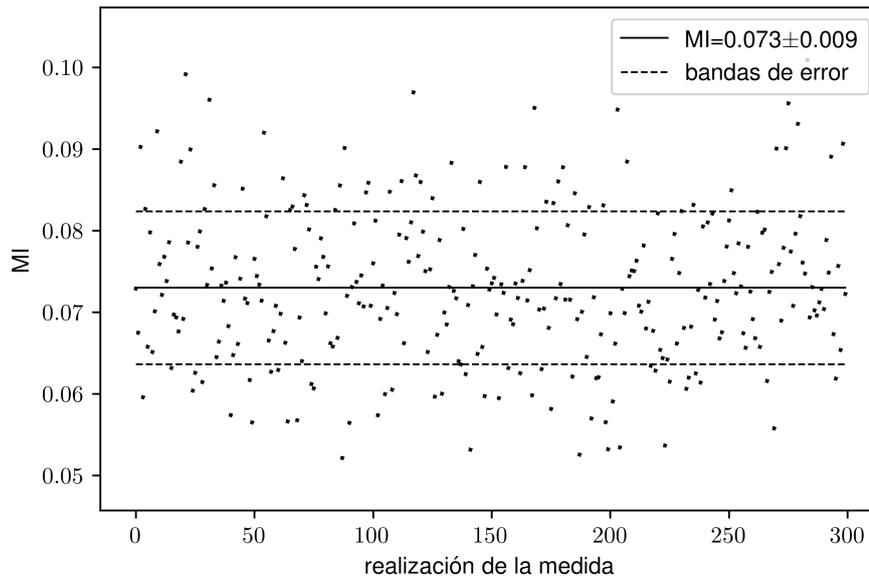


Figura 3.6: 300 realizaciones de $\mathbf{I}(X_{20}^t, X_{19}^t)$ con el método de KSG.

Se realizó el mismo experimento que en la sección 3.1 para detectar correlaciones entre escalas el Lorenz-96 de dos escalas, calculando la MI con el método de KSG. Los valores numéricos de $\mathbf{I}(X_4, Y_i)$ en función de i , con $i = 1, \dots, 256$ se muestran en la figura 3.7.

Al igual que con ASO (figura 3.4), se detectan las correlaciones con las distintas varia-

bles de la escala pequeña, con la diferencia de que las 32 variables de la escala pequeña acopladas con X_3 están más correlacionadas con X_4 que las 32 variables acopladas con X_5 , contrario a lo encontrado con el método ASO.

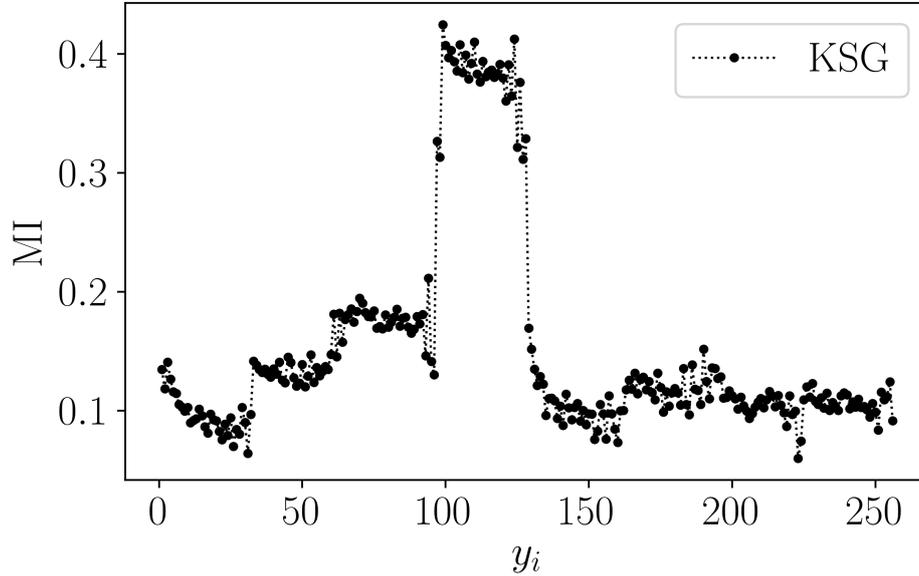


Figura 3.7: información mutua entre la variable X_4 de la gran escala y todas las 256 variables Y_j de la pequeña escala, calculada con el método de KSG.

3.3. Redes neuronales

Antes de repetir los experimentos anteriores con el método MINE, se analizó la convergencia del algoritmo en el Lorenz-96 de dos escalas. Se graficó la estimación de $\mathbf{I}(X_4, Y_{16})$ y $\mathbf{I}(X_4, Y_{128})$ (es decir, la MI entre una variable de la gran escala con una variable de la pequeña escala desacoplada y otra acoplada) en función de la época de entrenamiento. Recordar que minimizar la función de costo de la red neuronal es maximizar el valor de la información mutua (ecuación 1.27). Los valores numéricos de la MI se muestran en la figura 3.8. Se encuentra que el método converge a un rango de valores de MI fijo en cada caso.

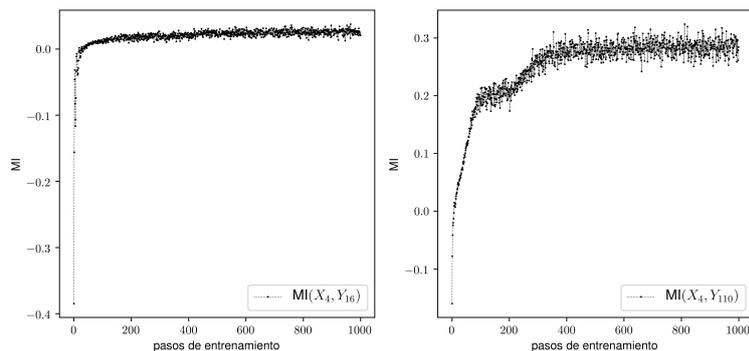


Figura 3.8: información mutua entre la variable X_4 de la gran escala y Y_{16} (derecha) y Y_{128} (derecha) de la pequeña escala.

Una vez mostrada la convergencia del algoritmo, se calculó la información mutua entre la variable X_{20} a tiempo presente y el resto de las variables X_i , $i = 1, \dots, 40$ a tiempo pasado para los atrasos temporales $\tau = 0, 5$ y 10 en el Lorenz-96 de una escala. Cada época de entrenamiento tiene como entrada series temporales de largo $n_{dat} = 5000$ puntos, en un total de $n_{ep} = 1000$ épocas. Dado que en la figura 3.8 se observa convergencia del modelo a partir de la época 500, se tomó como valor de MI al promedio de la estimación entre las épocas 500 y 1000. Los valores numéricos de $\mathbf{I}(X_{20}^t, X_i^{t-\tau})$ con $\tau = 0, 5$ y 10 muestran en las figuras 3.9 a, b y c respectivamente.

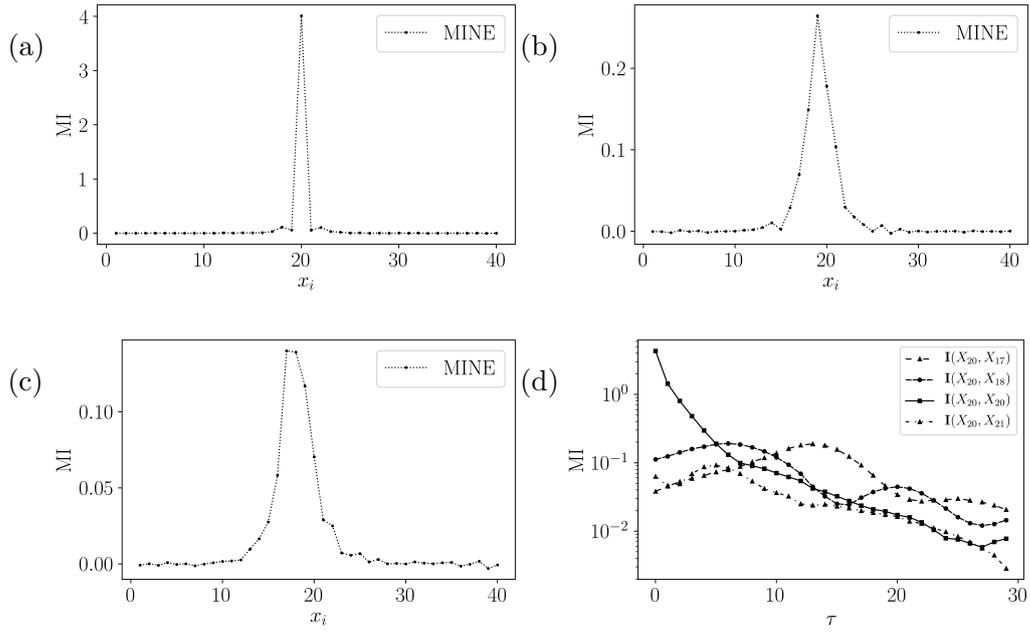


Figura 3.9: Información mutua con el método MINE de la variable X_{20} con todas las variables X_i del Lorenz 96 de una escala para (a) $\tau = 0$, (b) $\tau = 5$ y (c) $\tau = 10$. En (d) se muestra la información mutua de la variable X_{20} con X_{17} , X_{18} , X_{20} y X_{21} en función de τ .

Cuando no hay desfases temporales entre las series, el máximo de MI está en $i = 20$ (figura 3.9a). Cuando $\tau = 5$, el máximo de la MI ocurre en la variable X_{19} (figura 3.9b). A $\tau = 10$, es la variable X_{17} la que está correlacionada con el presente de X_{20} (figura 3.9c), mostrando claramente la direccionalidad del flujo de información desde las variables de la izquierda hacia las de la derecha.

Siguiendo la metodología de las secciones 3.1 y 3.2, se calculó $I(X_4, Y_i)$ en función de i , con $i = 1, \dots, 256$ en el modelo Lorenz-96 de dos escalas para estimar la utilidad del método MINE en la detección de correlaciones entre escalas. Los valores numéricos de la MI se muestran en la figura 3.10.

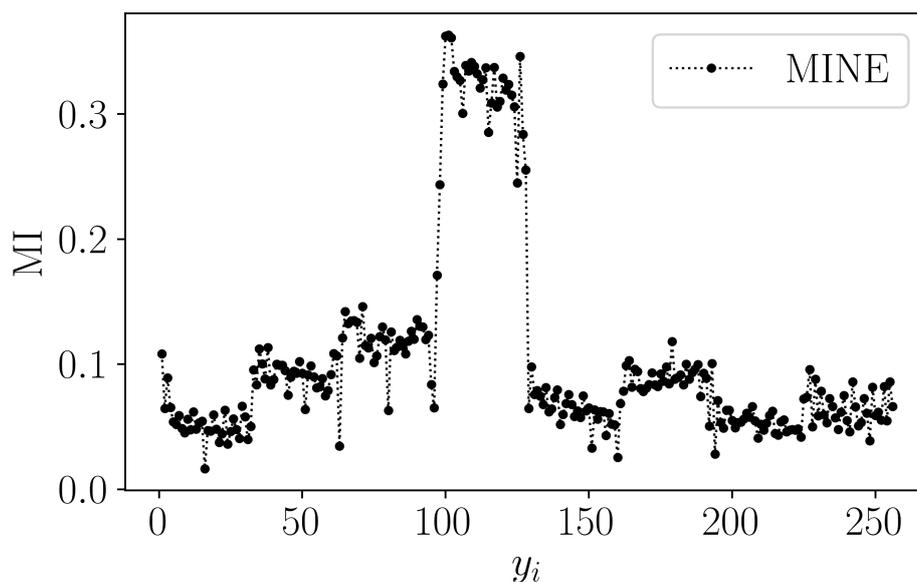


Figura 3.10: información mutua entre la variable X_4 de la gran escala y todas las 256 variables Y_j de la pequeña escala, calculada con el método de MINE.

Se observa que el método MINE puede capturar las interacciones entre la pequeña escala y la gran escala, siendo capaz de distinguir las variables Y_i , $i = 97, \dots, 128$ que interactúan directamente con la variable X_4 .

3.4. Comparación entre estimadores

En esta sección se comparan los tres estimadores de información mutua utilizados en las secciones 3.1, 3.2 y 3.3. En las figuras 3.11 a, b y c se muestran los valores de la información mutua, calculada con los tres estimadores, entre las series de tiempo de X_{20} a tiempo presente y el resto de las variables X_i , $i = 1, \dots, 40$ del Lorenz-96 de una escala para los tiempos pasados $\tau = 0, 5$ y 10 . En todos los casos, el método KSG dio los mayores valores de MI, seguido del Método MINE, y luego el ASO. Se observa claramente la similitud entre los métodos de KSG y MINE, detectando los mismos máximos de MI cuando existen desfases temporales en las series temporales (figuras 3.11b y 3.11c), aunque claramente discrepan en la magnitud de la MI cuando se compara X_{20} consigo misma en la figura 3.11a a tiempo presente. Cuando existen desfases temporales (3.11b y 3.11c), las medidas de MI estimadas con el análisis simbólico ordinal predicen correlaciones a tiempos pasados muy distintas a las medidas KSG y MINE: a $\tau = 5$ el pasado de X_{20} sigue teniendo más correlación sobre el presente de X_{20} que las variables vecinas, mientras que a $\tau = 10$ aparecen dos máximos de MI: uno en X_{17} (también observado con MINE y KSG), y otro en la misma variable X_{20} .

Las diferencias entre los estimadores también se observan cuando se utilizan series temporales del modelo Lorenz-96 de dos escalas: Los valores de $I(X_4, Y_i)$ para $i = 97, \dots, 128$ (MI entre la variable X_4 con las 32 variables de pequeña escala de su grupo local) son muy similares entre las medidas MINE y ASO, pero más grandes con el método KSG. Para el resto de las variables, el valor de la MI estimada con ASO es mucho mayor que con MINE y KSG.

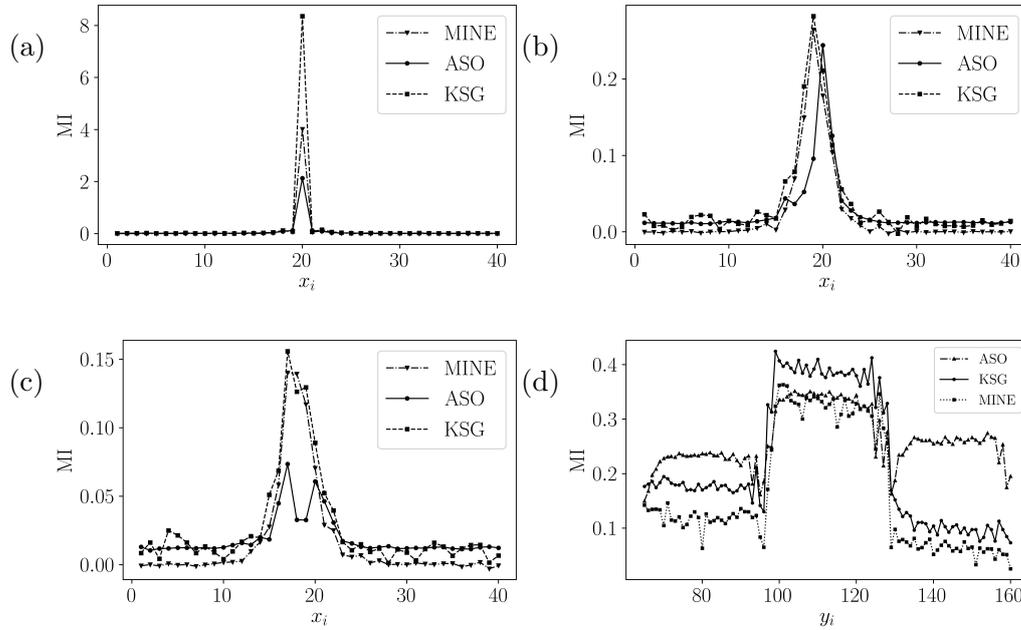


Figura 3.11: Estimación de $I(X_{20}, X_i)$ con los tres métodos cuando se atrasa en τ a la variable X_i . Arriba, izquierda: $\tau = 0$. Arriba, derecha: $\tau = 5$. abajo: $\tau = 10$.

La diferencia en la predicción de correlaciones del ASO respecto de KSG y MINE (figuras 3.11b y 3.11c) puede deberse a la diferencia conceptual de los métodos: el ASO es básicamente un análisis de la concavidad y conveccidad de las series temporales independiente de la magnitud de las series, en el sentido de que, mediante el uso de patrones, cuantifica las veces que la serie crece, decrece, oscila de determinada manera, etc., mientras que el método de KSG y MINE hacen un análisis más cuantitativo, el primero siendo un método determinista basado en la búsqueda de primeros vecinos, y el segundo un método estocástico basado en un problema de optimización.

4 Experimentos con el GOY

En este capítulo se analizan los estimadores de información mutua y transferencia de entropía en series de tiempo del sistema dinámico GOY. El objetivo es detectar la cascada directa de energía desde los modos altos a los modos bajos, manifestada a través del flujo de información en las series de tiempo de la energía desde los números de onda bajos hacia los altos. Para eso, se calcularán las medidas de información ya mencionadas entre el modo 9 de la energía y los modos 7, 8, 10, 11 a tiempos pasados.

4.1. Análisis simbólico Ordinal

En esta sección se aplican las medidas de información mediante el uso del análisis simbólico ordinal. El largo del vector ordinal es $D = 5$, y el largo de las series temporales es de $n_{dat} = 50,000$ puntos. En la figura 4.1a se muestran los valores numéricos de la información mutua entre el modo 9 de la energía y sus cuatro modos vecinos más cercanos a tiempos pasados, en función del lag temporal τ , medido en tiempos de observación ($t_{obs} = 0,1$, siendo $h = 0,001$ el paso de resolución del sistema). Se observa que a tiempos pasados el modo 9 está muy correlacionado a tiempos muy cortos ($\tau = 3$) con el modo 8, mientras que la correlación con el modo 10 es rápidamente decreciente, al igual que el modo 11. La interacción con el modo 7 comienza siendo pequeña a comparación con los otros, pero se sostiene a lo largo del tiempo, siendo la interacción predominante a partir de $\tau = 15$. En la figura 4.1 b se muestran los valores numéricos de la transferencia de entropía hacia el modo 9 desde sus cuatro primeros vecinos.

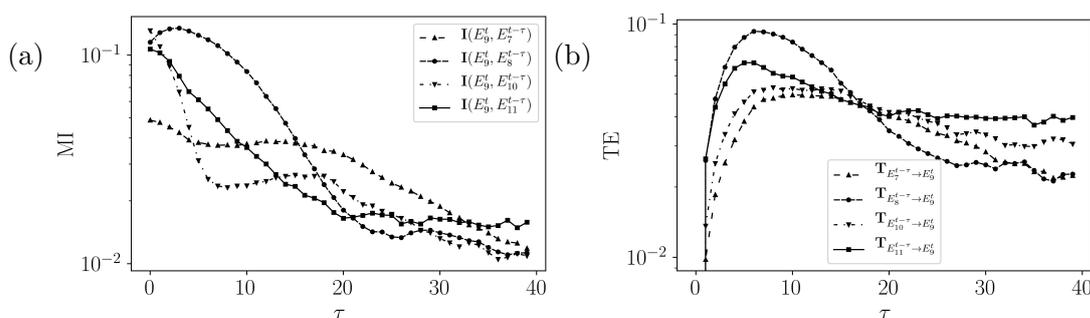


Figura 4.1: (a): $I(E_9^t, E_i^{t-\tau})$ para varios τ , con $i = 7, 8, 10, 11$. (b): $T_{E_i^{t-\tau} \rightarrow E_9^t}$ para varios τ , con $i = 7, 8, 10, 11$, calculadas con el ASO.

La transferencia de entropía muestra interacciones entre los modos distinta a la información mutua: La mayor correlación en la TE se da entre el modo 8 y el 9, como

es de esperar, sólo que el máximo ocurre a tiempos pasados mayores ($\tau = 6$) que en la información mutua ($\tau = 3$). Inesperadamente, la segunda mayor interacción a tiempos pasados proviene del modo 11, con tiempos de correlación menores ($\tau = 4$) a los tiempos de correlación del modo 8, mientras que desde el modo 7 y 10 la interacción es menor a la de los modos 8 y 11.

El error al estimar la MI con el ASO se lo calculó haciendo 300 realizaciones de la medida entre el modo 10 y el modo 9 a tiempo presente sobre series temporales independientes (figura 4.2a), mientras que el error de la TE (figura 4.2 b) se lo calculó haciendo 300 realizaciones de la medida desde el modo 8 hasta el modo 9 para un tiempo de atraso temporal $\tau = t_{obs}$. Los errores estadísticos asociados son de 0,9% para la MI y 0,3% para la TE.

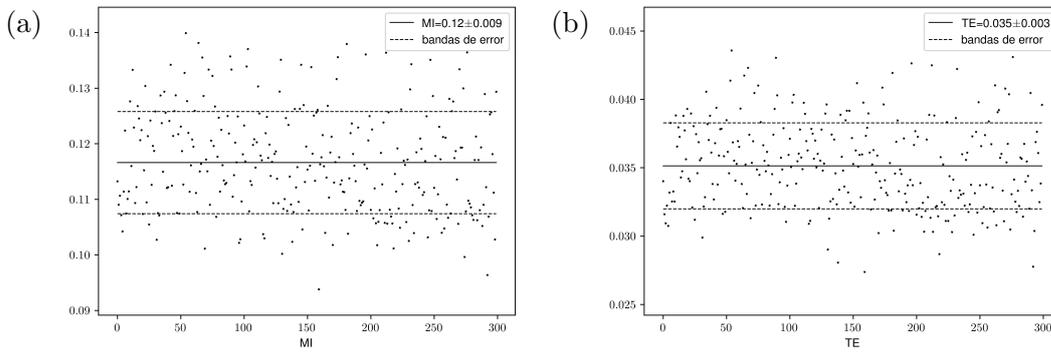


Figura 4.2: 300 realizaciones de (a): $\mathbf{I}(E_{10}, E_9)$ y (b): $\mathbf{T}_{E_9^{t-\tau} \rightarrow E_{10}^t}$ para $\tau = t_{obs}$, calculadas con el ASO, con sus medias y desviaciones estándar.

4.2. Método de KSG

En esta sección se aplican las medidas de información mediante el uso del método de KSG. El número de primeros vecinos del algoritmo es $k = 3$, y el largo de las series temporales es $n_{dat} = 15,000$. En la figura 4.3 se muestran los valores numéricos de MI entre el modo 9 de la energía a tiempo presente y los modos 7, 8, 10, 11 de la energía a tiempo pasado, en función del atraso temporal τ .

Se observa que los modos 7 y 8 a tiempo pasado claramente influyen más en la dinámica del modo 9 que los modos 10 y 11. Los tiempos característicos de la correlación de los modos 7 y 8 son muy distintos: $\mathbf{I}(E_9^t, E_8^{t-\tau})$ es máxima a $\tau = 5$, mientras que $\mathbf{I}(E_9^t, E_7^{t-\tau})$ es máxima a $\tau = 16$. Las observaciones de la direccionalidad de la MI indican que el método de KSG detecta la cascada directa de energía, dado que sólo se muestra interacciones provenientes desde los modos pequeños hacia los grandes.

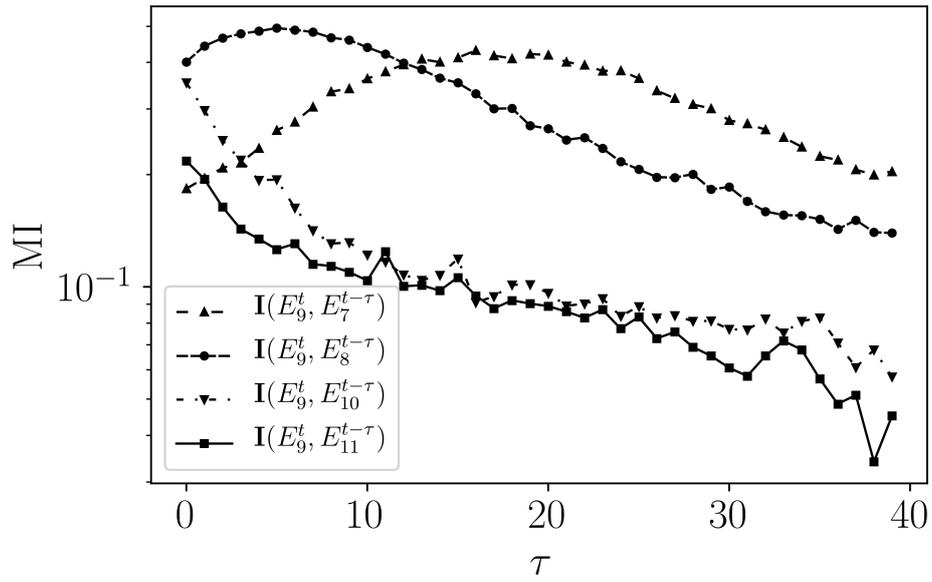


Figura 4.3: $I(E_9^t, E_i^{t-\tau})$ para varios τ , con $i = 7, 8, 10, 11$, calculadaa con el método de KSG.

El error al estimar la MI con el método de KSG se la calculó haciendo 300 realizaciones de la medida entre el modo 10 y el modo 9 a tiempo presente sobre series temporales independientes (figura 4.4) El error asociado es de 1 %.

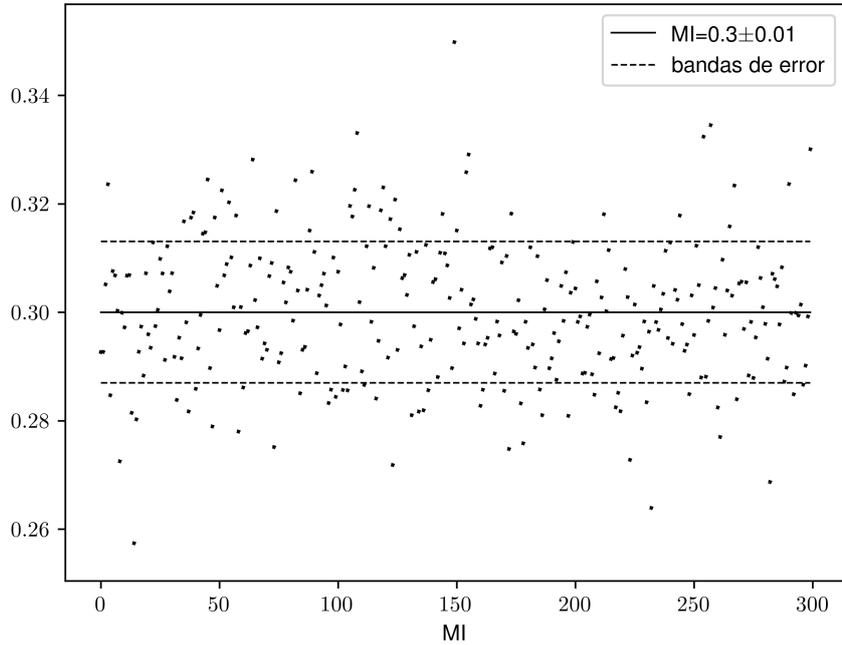


Figura 4.4: 300 realizaciones de $\mathbf{I}(E_{10}, E_9)$ calculada con el método de KSG, con su media y desviación estándar.

4.3. Redes Neuronales

En primera medida, se analizó la convergencia de MINE aplicado a las series temporales de la energía en el modelo GOY. Para eso, se calculó la información mutua entre los modos 8 y 9 de la energía a tiempo presente. Se entrenó la red neuronal durante 1000 épocas con series temporales de largo $n_{dat} = 5000$. El resultado se muestra en la figura 4.5a. Se obtiene que el método converge a partir de aproximadamente la época 400 en un rango fijo de MI. El valor de la medida es el valor medio de la MI entre las épocas 400 y 1000, y el error de la medida es su desviación estándar (figura 4.5 b). Se obtiene que el error de la estimación es 4%.

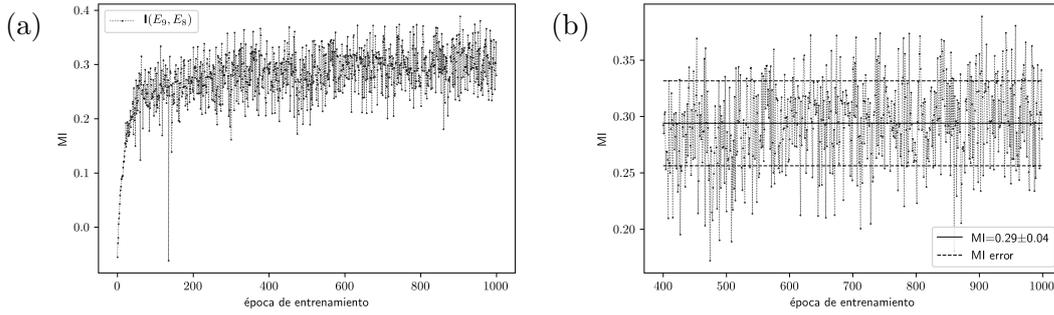


Figura 4.5: (a): $\mathbf{I}(E_9, E_8)$ calculada con el método de MINE para 1000 épocas de entrenamiento. (b): Promedio y desviación estándar de la MI entre la época 400 y la época 1000, una vez que converge la medida.

Corroborada la convergencia de la estimación, se calculó la MI mediante el uso del método MINE, durante 1000 épocas de entrenamiento, entre el modo 9 de la energía a tiempo presente y los modos 7, 8, 10 y 11 de la energía a tiempo pasado, en función del atraso temporal τ . Cada punto de la figura 4.6 es el promedio de la MI entre las épocas 400 y 1000.

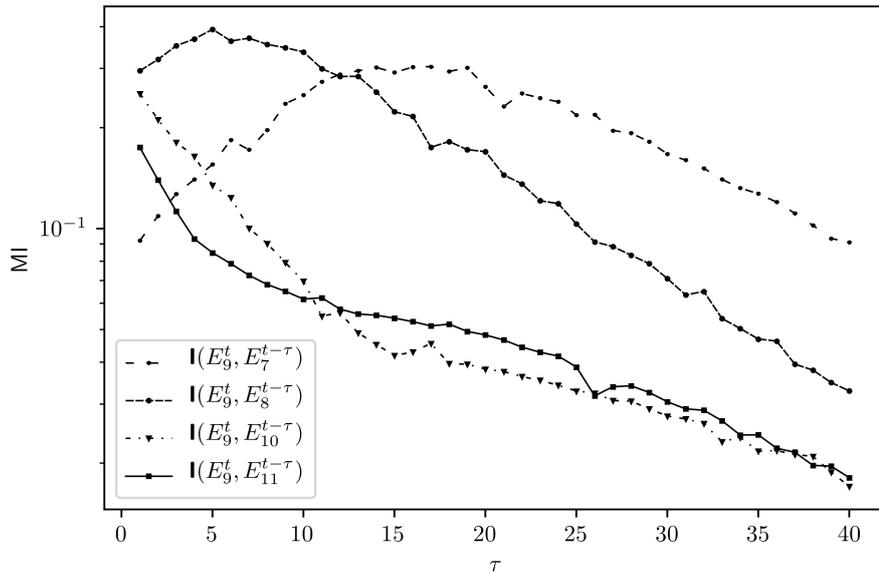


Figura 4.6: $\mathbf{I}(E_9^t, E_i^{t-\tau})$ para varios τ , con $i = 7, 8, 10, 11$, calculada con el método de MINE.

La correlación de los modos 10 y 11 a tiempo pasado con el presente del modo 9 de la energía disminuyen con el atraso temporal, mientras que el modo 8 presenta el máximo de correlación en $\tau = 5$ y el modo 7 en $\tau = 17$. Este resultado muestra la clara

direccionalidad de la información en el modelo GOY, mostrando la cascada de energía desde los números de onda bajos a los altos, observada indirectamente con el flujo de la MI calculada con el método MINE.

4.4. Comparación entre estimadores

En esta sección se comparan los tres estimadores de información mutua utilizados en las secciones 4.1, 4.2 y 4.3. En las figuras 4.7 a, b, c y d se muestran los valores de la información mutua, calculada con los tres estimadores, entre las series de tiempo de E_9 a tiempo presente y las series de tiempo de E_7 , E_8 , E_{10} y E_{11} a tiempo pasado, respectivamente. En todos los casos, El valor de la MI es más grande cuando se la estima con el método de KSG, seguido por el MINE y luego el ASO.

Tanto KSG como MINE detectan correlación del modo E_9 con el modo E_7 a tiempo pasado (figura 4.7a), pero con distintos máximos de correlación: $\tau = 15 - 18$ en el método MINE y $16 - 19$ en el método de KSG. Esta direccionalidad de la información no se detecta con el ASO.

No es el caso cuando se calcula $\mathbf{I}(E_9^t, E_8^{t-\tau})$ (figura 4.7b): todas las medidas capturan la marcada correlación entre el pasado de E_8 y el presente de E_9 . El ASO presenta el máximo de la MI en $\tau = 4$, el método KSG en $\tau = 6$ y el MINE en $\tau = 5$.

Por otro lado, ningún método (Excepto por la TE, ver figura 4.1b) detectó flujos de información desde los modos altos a los modos bajos (figuras 4.7c y d).

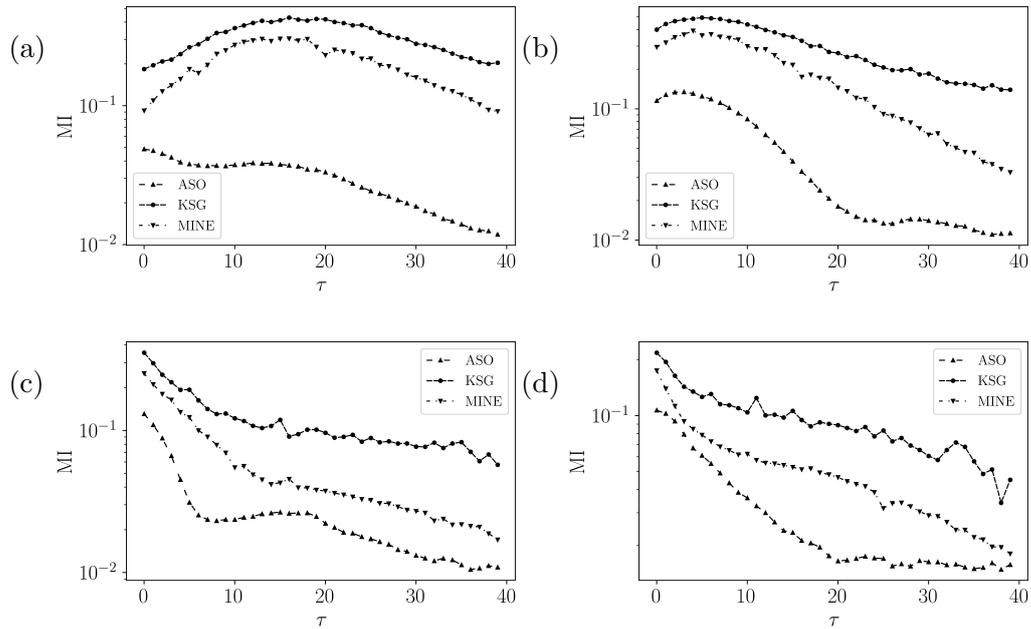


Figura 4.7: Estimación de la MI entre el presente de E_9 y el pasado de sus cuatro primeros vecinos con los tres métodos, en función del atraso temporal τ : (a) $\mathbf{I}(E_9^t, E_7^{t-\tau})$, (b) $\mathbf{I}(E_9^t, E_8^{t-\tau})$, (c) $\mathbf{I}(E_9^t, E_{10}^{t-\tau})$, (d) $\mathbf{I}(E_9^t, E_{11}^{t-\tau})$.

Se concluye que los tres estimadores de MI resultan eficaces a la hora de detectar, de forma indirecta, la cascada directa de energía en el modelo sintético de turbulencia GOY.

5 Experimentos con la ecuación barotrópica de la vorticidad

En el capítulo 4 se aplicaron las medidas de información al modelo GOY, un sistema dinámico modelado por el sistema de ecuaciones ordinarias 2.4 diseñado para imitar el espectro de energía propio de un flujo turbulento en tres dimensiones, cuyas interacciones entre nodos son conocidas. En esta sección, se aplican los cuantificadores de información mutua y transferencia de entropía a las series temporales de la energía y enstrofia, calculadas a partir de las soluciones de las ecuaciones 2.8. El objetivo es observar de forma indirecta la cascada de enstrofia desde los números de onda bajos a los números de onda altos en el modelo utilizando las medidas de información.

Los experimentos de esta sección son similares a los realizados con el modelo GOY en el capítulo 4: a partir de las series de tiempo de la energía y enstrofia, se pretende observar indirectamente las cascadas de energía y enstrofia en el sistema de turbulencia bidimensional. Para eso, se calcula la información mutua entre un modo de la energía (S_i) y de la enstrofia (Z_i) a tiempo presente con los cuatro modos más cercanos a tiempos pasados, así como la transferencia de entropía de esos vecinos hacia S_i y Z_i . En el rango inercial de la cascada de enstrofia (desde los números de onda del forzado hacia los números de onda altos), en principio, se debería observar un flujo de información en las series de la enstrofia hacia los números de onda altos, y ningún flujo de información en las series temporales de la energía [32]. En el rango inercial de la cascada de energía (desde los números de onda del forzado hacia los números de onda bajos), por otro lado, se debería observar un flujo de información en las series de la energía hacia los números de onda bajos, y ningún flujo de información en las series temporales de la enstrofia [32].

En todos los casos en que se calcula la información mutua o transferencia de entropía entre series de tiempo atrasadas, el retardo temporal está medido en tiempos de observación ($\tau = t_{obs} = 0,25$).

No se realizaron experimentos con el método MINE porque las series temporales del modelo barotrópico no son estacionarias (ver figura 2.8), a diferencia de las series temporales del Lorenz-96 (figura 2.2) y el modelo GOY (figura 2.4). Para la red neuronal esto es perjudicial porque la distribución de probabilidad asociada a las variables de los modelos no es estacionaria. Eso hace que lo incorporado en una época de entrenamiento no le sirva para la próxima, resultando en la no convergencia del método.

5.1. Análisis simbólico ordinal

En esta sección se aplican las medidas de información mediante el uso del análisis simbólico ordinal. El largo del vector ordinal es $D = 5$, y el largo de las series temporales

es de $n_{dat} = 50,000$ puntos.

En la figura 5.1a se muestran los valores numéricos de la MI entre las series temporales de la enstrofia del modo 25 con sus cuatro vecinos más cercanos a tiempo pasado, estimada con ASO, en función de τ . El modo 23 es el que más correlación tiene con el modo 25, aunque esta interacción siempre es decreciente. Lo mismo sucede con el modo 26. La MI entre Z_{25} con el pasado de Z_{24} , en cambio, es máxima cuando el atraso temporal es $\tau = 3$. En la figura 5.1b se muestran los valores numéricos de la TE hacia el modo 25 de la enstrofia, proveniente de sus cuatro vecinos más cercanos. El máximo valor de la TE se da en el modo 25, y en menor medida desde el modo 24, cuando $\tau = 3$. La TE desde los modos 26 y 27 hacia el modo 25 es menor, comparada con la de los modos 23 y 24, y presentan su máximo valor cuando $\tau = 2$. Estos resultados muestran una mayor correlación entre el modo 25 de la enstrofia a tiempo presente con los modos menores, mostrando cierta direccionalidad del flujo de información hacia los números de onda altos, correspondientes a la cascada de enstrofia.

Algo inesperado ocurre cuando se calcula la MI entre las series temporales de la energía del modo 25 con sus cuatro vecinos más cercanos a tiempo pasado (figura 5.1c). Tanto las interacciones como los valores de la MI son muy similares a los resultados obtenidos con la enstrofia (figura 5.1a). Lo mismo ocurre con el análisis de la TE hacia el modo 25 de la energía (figura 5.1d): Se observa una transferencia de información desde los modos bajos a los modos altos en las series de tiempo de S . Este resultado no es el esperado, ya que en el régimen inercial de la cascada de enstrofia no hay transferencias de energía.

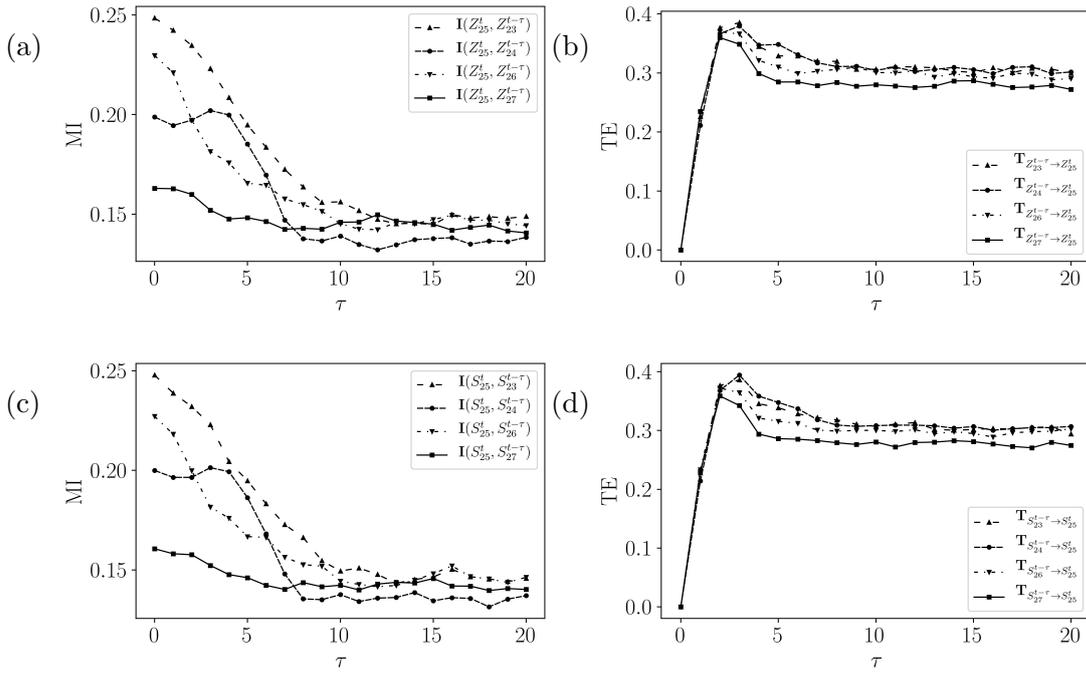


Figura 5.1: (a) Información mutua entre el modo 25 de la enstrofia y sus cuatro primeros vecinos en función del lag temporal τ , (b) transferencia de entropía desde los cuatro primeros vecinos hacia el modo 25 de la enstrofia, (c) información mutua calculada con el ASO entre el modo 25 de la energía y sus cuatro primeros vecinos en función del lag temporal τ , (d) transferencia de entropía desde los cuatro primeros vecinos hacia el modo 25 de la energía, todas calculadas con ASO.

En la figura 5.2 (a) se muestra la MI entre la serie de tiempo de la enstrofia del modo 5 a tiempo presente y la de los modos 3,4, 6 y 7 a tiempo pasado, estimada con ASO, en función del atraso temporal τ . Ningún modo de la enstrofia muestra un máximo de MI a tiempos pasados, y la correlación disminuye con el tiempo. En la figura 5.1 (b) se muestra la transferencia de entropía desde los modos 3, 4, 6 y 7 hacia el modo 5 de la enstrofia. No se observa direccionalidad alguna de información, ya que todos los máximos de la TE calculada ocurren en $\tau = 4$, y los mayores valores de TE son los correspondientes a los modos 3 y 6. Este es el resultado esperado, ya que no se deberían observar flujos de enstrofia en la zona del espectro donde ocurren las cascadas de energía.

En la figura 5.2 (c) se muestra la MI entre la serie de tiempo de la energía del modo 5 a tiempo presente y la de los modos 3,4, 6 y 7 a tiempo pasado, estimada con ASO, en función del atraso temporal τ . De forma similar a la enstrofia, ningún modo muestra un máximo de MI a tiempos pasados. En la figura 5.1 (d) se muestra la transferencia de entropía desde los modos 3, 4, 6 y 7 hacia el modo 5 de la energía. Al igual que con la enstrofia, no se observan direccionalidades de la información. Esto significa que no se observa la cascada inversa de energía.

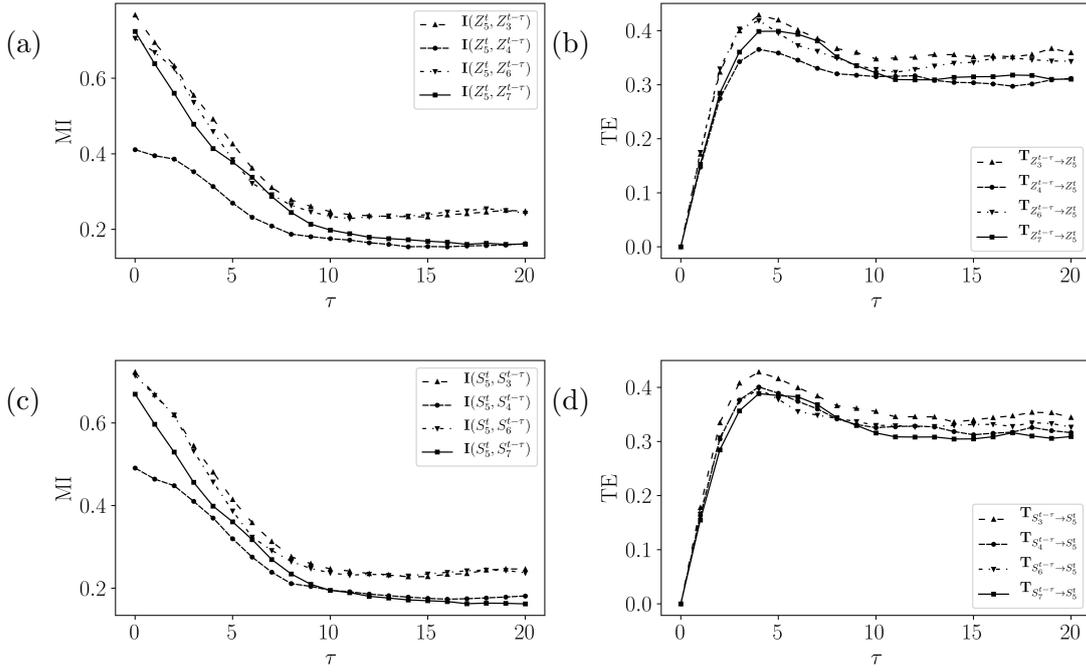


Figura 5.2: (a) Información mutua entre el modo 5 de la enstrofía y sus cuatro primeros vecinos en función del lag temporal τ , (b) transferencia de entropía desde los cuatro primeros vecinos hacia el modo 5 de la enstrofía, (c) información mutua calculada con el ASO entre el modo 5 de la energía y sus cuatro primeros vecinos en función del lag temporal τ , (d) transferencia de entropía desde los cuatro primeros vecinos hacia el modo 5 de la energía, todas calculadas con ASO.

5.2. Método de KSG

En la figura 5.3a se muestran los valores numéricos de la MI entre las series temporales de la enstrofía del modo 25 con sus cuatro vecinos más cercanos en función del lag temporal τ , estimada con el método de KSG. Claramente se observa un flujo de información desde el modo 24 hacia el 25 con un máximo en $\tau = 2$. También se observa, en menor medida, un flujo de información desde el modo 23 hacia el 25, con un máximo en $\tau = 7$. No se observan correlaciones a tiempos pasados de los modos 26 y 27 con el 25. El método de KSG aplicado a las series temporales de la enstrofía muestra una clara transferencia de información desde los números de onda bajos a los altos, pudiendo captar la cascada directa de enstrofía.

En la figura 5.3b se muestran los valores numéricos de la MI entre las series temporales de la energía del modo 25 con sus cuatro vecinos más cercanos en función del lag temporal τ , estimada con el método de KSG. Lo que ocurre en la sección 5.1 ocurre también con el método de KSG: los valores de MI en las series de la energía son muy similares, tanto en valor como en la direccionalidad del flujo de información, a los valores de MI de las series de la enstrofía, mostrando una cascada directa de energía que no se esperaba observar.

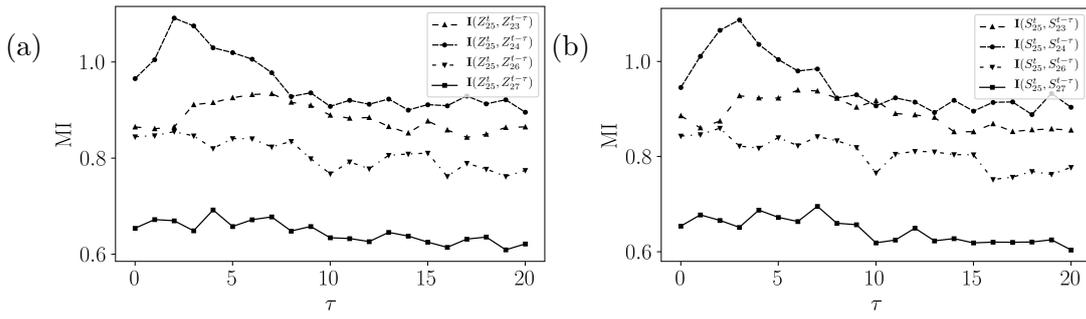


Figura 5.3: Información mutua entre el modo 25 y sus cuatro primeros vecinos en función del lag temporal τ calculada con el método de KSG para con series temporales de (a) la enstrofia y (b) la energía.

En las figuras 5.4a (b) se muestran los valores numéricos de la MI entre las series temporales de la enstrofia (energía) del modo 5 con sus cuatro vecinos más cercanos en función del lag temporal τ , estimada con el método de KSG. No se observan correlaciones algunas entre el modo 5 de la energía y enstrofia. Los valores de MI mostrados en las figuras son propios de series temporales independientes con muy poca correlación entre sí. El método de KSG no es concluyente a la hora de observar la cascada inversa de energía (desde los números de onda altos a los bajos).

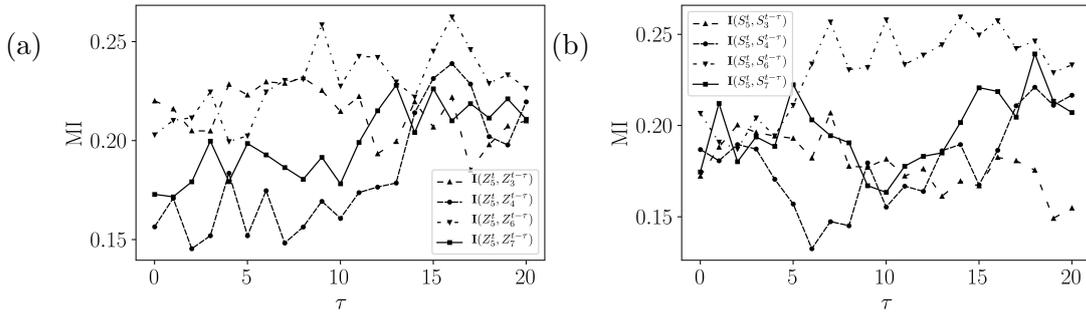


Figura 5.4: Información mutua entre el modo 5 y sus cuatro primeros vecinos en función del lag temporal τ calculada con el método de KSG para con series temporales de (a) la enstrofia y (b) la energía.

6 Conclusiones

En este trabajo se utilizaron como medidas de correlación a la información mutua y la transferencia de entropía, estimándolas con tres métodos distintos. Las medidas resultaron útiles para detectar causalidades implícitas entre variables que no están modeladas en las ecuaciones del sistema dinámico Lorenz-96, así como las interacciones entre escalas en el Lorenz-96 de dos escalas.

Las medidas resultaron eficientes en el modelo GOY para detectar el flujo de información en las series de tiempo de la energía desde los números de onda altos a los números de onda bajos, observando de forma indirecta con medidas de información la cascada directa de energía en un modelo sintético de turbulencia, en concordancia con los resultados de [13].

Se logró detectar de forma indirecta, utilizando las medidas de información propuestas, la cascada directa de enstrofia en el modelo barotrópico, aunque se observó una cascada de energía hacia los números de onda altos en este modelo de turbulencia bidimensional cuando no se la debería ver. No se observó la cascada inversa de energía.

Se logró aplicar exitosamente el método MINE en series temporales, siendo los resultados muy similares a los del método de KSG. El método MINE es rápido y no es computacionalmente demandante. La desventaja es el requerimiento de muchos datasets para entrenar la red neuronal.

El ASO, por otro lado, es un método simple y eficiente, aunque en este trabajo se observan discrepancias con respecto al MINE y KSG. Esto puede deberse a la diferencia conceptual de los métodos: el ASO es básicamente un análisis de la concavidad y convexidad de las series temporales independiente de la magnitud de las series, en el sentido de que, mediante el uso de patrones, cuantifica las veces que la serie crece, decrece, oscila de determinada manera, etc., mientras que el método de KSG y MINE hacen un análisis cuantitativo, el primero siendo un método determinista basado en la búsqueda de primeros vecinos, y el segundo un método estocástico basado en un problema de optimización.

Como perspectivas a futuro, se espera aplicar los cuantificadores de información a modelos numéricos utilizados para el pronóstico meteorológico, con el fin de medir el impacto del aumento de resolución y de diferentes parametrizaciones de subgrilla.

Bibliografía

- [1] RG Prinn. Development and application of earth system models. *Proceedings of the National Academy of Sciences*, 110:3673–3680, 2013.
- [2] A.A. Tsonis and P.J. Roebber. The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, 469:497 – 504, 2004.
- [3] SH Strogatz DJ Watts. Collective dynamics of “small-world” networks. *Nature*, 393:440–442, 1998.
- [4] CWJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 37(3):424–438, 1969.
- [5] EN Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–141, 1963.
- [6] CE Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
- [7] S Kullback. *Information Theory and Statistics*. Dover Publications, 31 East 2nd Street Mineola NY 11501-3852 USA, 1968.
- [8] T Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, 2000.
- [9] Milan Paluš, Vladimír Komárek, Zbyněk Hrnčíř, and Katalin Štěrbová. Synchronization as adjustment of information rates: Detection from bivariate time series. *Physical Review E*, 63(4):046211, 2001.
- [10] SL Lauritzen. *Graphical Models*. Oxford University Press, 116-117 High St, Oxford OX1 4BZ, United Kingdom, 2000.
- [11] N Marwan J Kurths J Runge, J Heitzig. Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy. *Physical Review E*, 86:061121, 2012.
- [12] K Ohkitani M Yamada. Spectrum of a chaotic model of three-dimensional turbulence. *Journal of the Physical Society of Japan*, 56(12):4210–4213, 1987.
- [13] N Smith R De Marco M Materassi, G Consolini. Information theory analysis of cascading process in a synthetic model of fluid turbulence. *Entropy*, 16(3):1272–1286, 2013.

- [14] OA Rosso M Pulido. Model selection: Using information measures from ordinal symbolic analysis to select model subgrid-scale parameterizations. *Journal of the Atmospheric Sciences*, 74(10):3253–3269, 2017.
- [15] EN Lorenz. Predictability - a problem partly solved. *Proc. Seminar on Predictability, Shinfield Park, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasting*, pages 1–81, 1996.
- [16] X Calbet R Lopez-Ruiz, HL Mancini. A statistical measure of complexity. *Physics letters A*, 209(5-6):321–326, 1995.
- [17] J Lin. Divergence measures based on the shannon entropy. *IEEE Trans. Inf. Theory*, 37(1):145–151, 1991.
- [18] W Just HG Schuster. *Deterministic chaos: An introduction*, pp. 320. 1995.
- [19] B Pompe C Bandt. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88(17):174102, 2002.
- [20] G Peter A Kraskov, H Stögbauer. Estimating mutual information. *Phys. Rev. E*, 69(6):066138, 2004.
- [21] GV Steeg. Non-parametric entropy estimation toolbox. <https://github.com/gregversteeg/NPEET>.
- [22] S Rajeswar S Ozair Y Bengio A Courville RD Hjelm MI Belghazi, A Baratin. Mutual information neural estimation. <https://arxiv.org/abs/1801.04062>, 2018.
- [23] JA Hertz. *Introduction to the theory of neural computation*. CRC Press, 6000 Broken Sound Parkway NW, Suite 300. Boca Raton, FL 33487-2742, 1999.
- [24] SRS Varadhan MD Donsker. Asymptotic evaluation of certain markov process expectations for large time. iv. *Communications on Pure and Applied Mathematics*, 36(2):183–212, 1983.
- [25] B Arindam. On bayesian bounds. *roceedings of the 23rd International Conference on Machine Learning*, pages 81–88, 2006.
- [26] M Zgubic. Information neural estimator implemented in tensorflow.
- [27] J Ba DP Kingma. Adam: A method for stochastic optimization. <https://arxiv.org/abs/1412.6980>.
- [28] H Abarbanel. *Analysis of Observed Chaotic Data*. Springer-Verlag, p. 272, 1996.
- [29] N Rana. Shell-turbulence. <https://github.com/navdeeprana/shell-turbulence>.
- [30] PC Matthews SM Cox. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, 2002.

- [31] R Fjørtoft. One the changes in spectral ditribution of kinetic energy for twodimensional, nondivergent flow. *Tellus*, 5(3):225–230, 1953.
- [32] D Dowling P Kundu, I Cohen. *Fluid mechanics. Ch. 14, p. 623-626*. Elsevier, 2000.
- [33] J Penn. Shallow water equations in python. <https://github.com/jamesp/shallowwater>.
- [34] CC Henning I. Marinov CY Tam IM Heldang K Vallis KS Smith, G Boccaletti. Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.*, 469:13–48, 2002.
- [35] JC McWilliams. The emergence of isolated coherent vortices in turbulent flow. *AIP Conference Proceedings*, 146:21–43, 1984.
- [36] C. Rocha F. Poulin M Jansen, R Abernathey. Python quasigeostrophic model. <https://pyqg.readthedocs.io/en/latest/>.