Análisis de Series de Tiempo por medio del uso de herramientas provenientes de la Teoría de la Información.

Diego Mateos. PhD

Instituto de Matemática Aplicada del Litoral

Universidad Autónoma de Entre Ríos

Resumen

- **Introducción a los conceptos de teoría de la información**
  Entropías (Shannon, Rényi, Tsallis, Fisher)
  Divergencias de (Kullbak-Leible y Jensen-Shannon)
  Complejidades (Estadística, Lempel-Ziv)

  Análisis de señales mediante planos complejidad - entropía :
  Estudio de pacientes epilepticos.
  Caracterización de la evolución clínica de tratamientos farmacológico.
  Caracterización de estadios de sueño.

Introducción a los conceptos básicos de la Teoría de la información.

Que es la Teoría de la Información ?

Es el estudio de las leyes que rigen la transmisión, almacenamiento y procesamiento de la información en un sistema. [Shannon 1948]

Aplicaciones

- Física.
- Ingeniería.
- Matemática.
- Biología.
- Lingüística.
- Neurociencias.

Shannon, Bell System Technical Journal. 1948

Por que utilizar la TI en el análisis de señales?

- Nos permite analizar y teorizar como manejan **información** los sistemas.

- En el análisis de señales, permite extraer información importante, que mediante los métodos clásicos no es posible.

- Posee un amplio abanico de posibilidades de implementación en diferentes áreas.

- Relativamente simples de aplicar, con alta velocidad de cómputo.

Entropías

Bits
'Information'
Messages
"Choice"
...001001011
Gatlin
Shannon
Wiener
Brillouin
0
1
"Uncertainty"
Printed Text
Data
ENTROPY

Entropía de Shannon (H)

*Es la cantidad de información que se desconoce del sistema. En otras palabras la **incerteza** que tenemos del mismo.*

Sistema de dos estados = Moneda (cara o seca)

Moneda normal

50 % cara
50 % seca

↓

Alta incerteza

↓

H max

Moneda mágica

100 % cara
0 % seca

↓

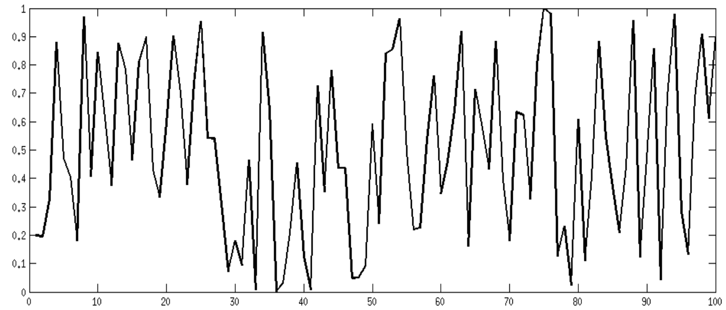
Baja incerteza

↓

H = 0

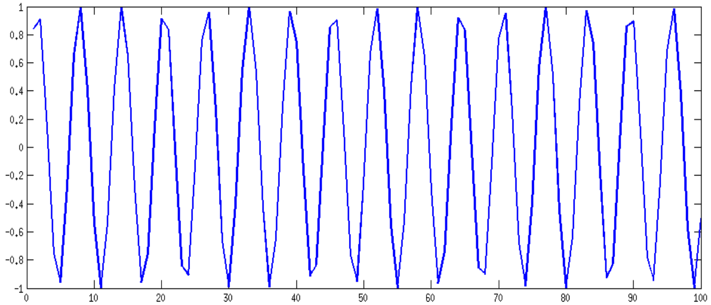
Entropía de Shannon (H)

Señales altamente fluctuante (ej. EEG en Vigilia)



H Alta

Señales con alta periodicidad (ej. Ataque epiléptico o Sueño profundo)



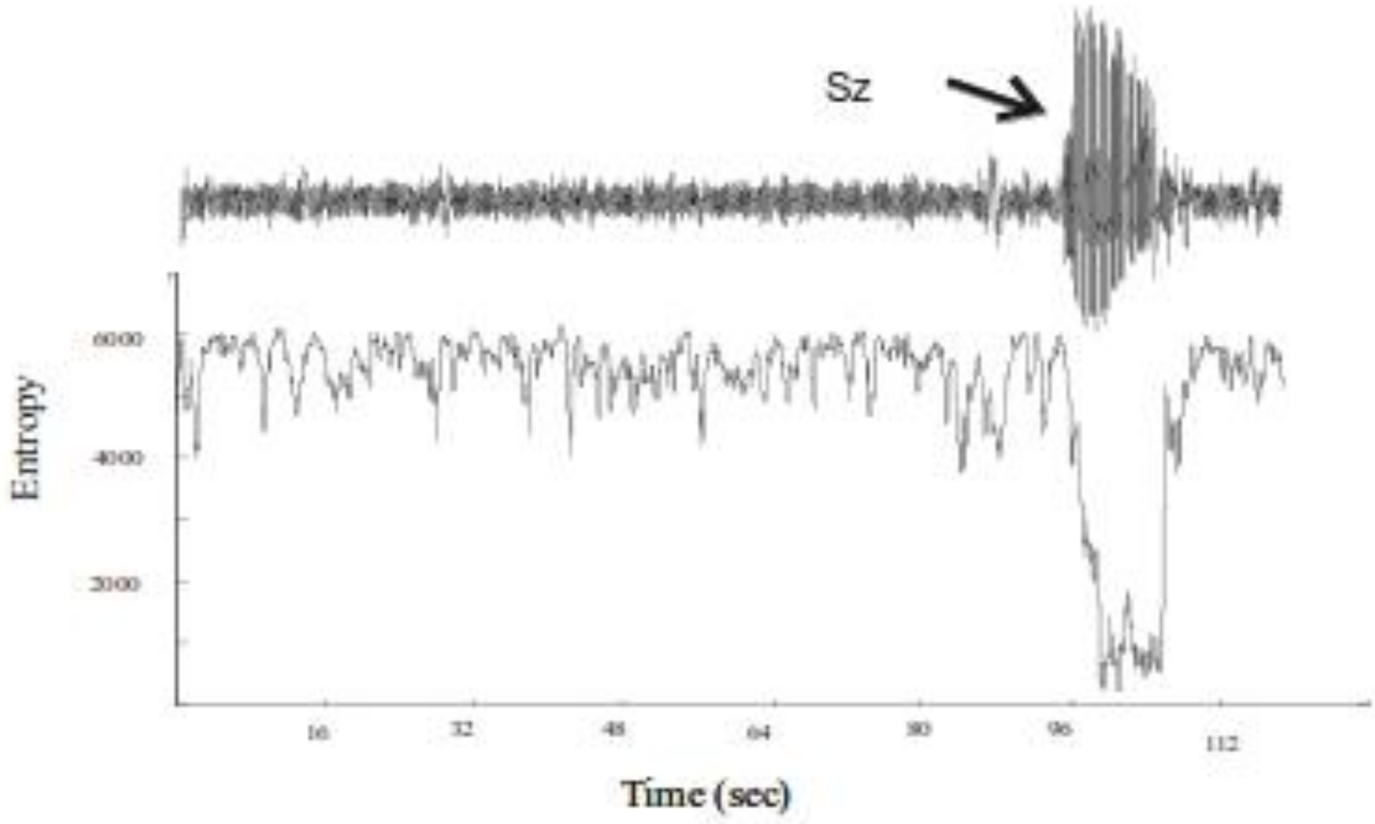
H Baja

Entropía de Shannon

Si tenemos una variable aleatoria X que puede tomar los valores $\{x_1,\ x_2, \ldots, x_n\}$ y la cual tiene una distribución de probabilidad asociada $P(X),\ P(x_i) = P_i$ se define la entropía de Shannon como :

$$H[P] = -\sum_{i=1} P_i . \log(P_i)$$

Ataque epiléptico en señales de MEG



Guevara, Mateos PRE. 2016

Existen otro tipos de entropías.

Las cuales pueden darnos otro tipo de información

Tsallis

$$S_q[P] = \frac{1}{q-1} \sum_{j=1}^{n} \left[P_j - (P_j)^q \right]$$

Rényi

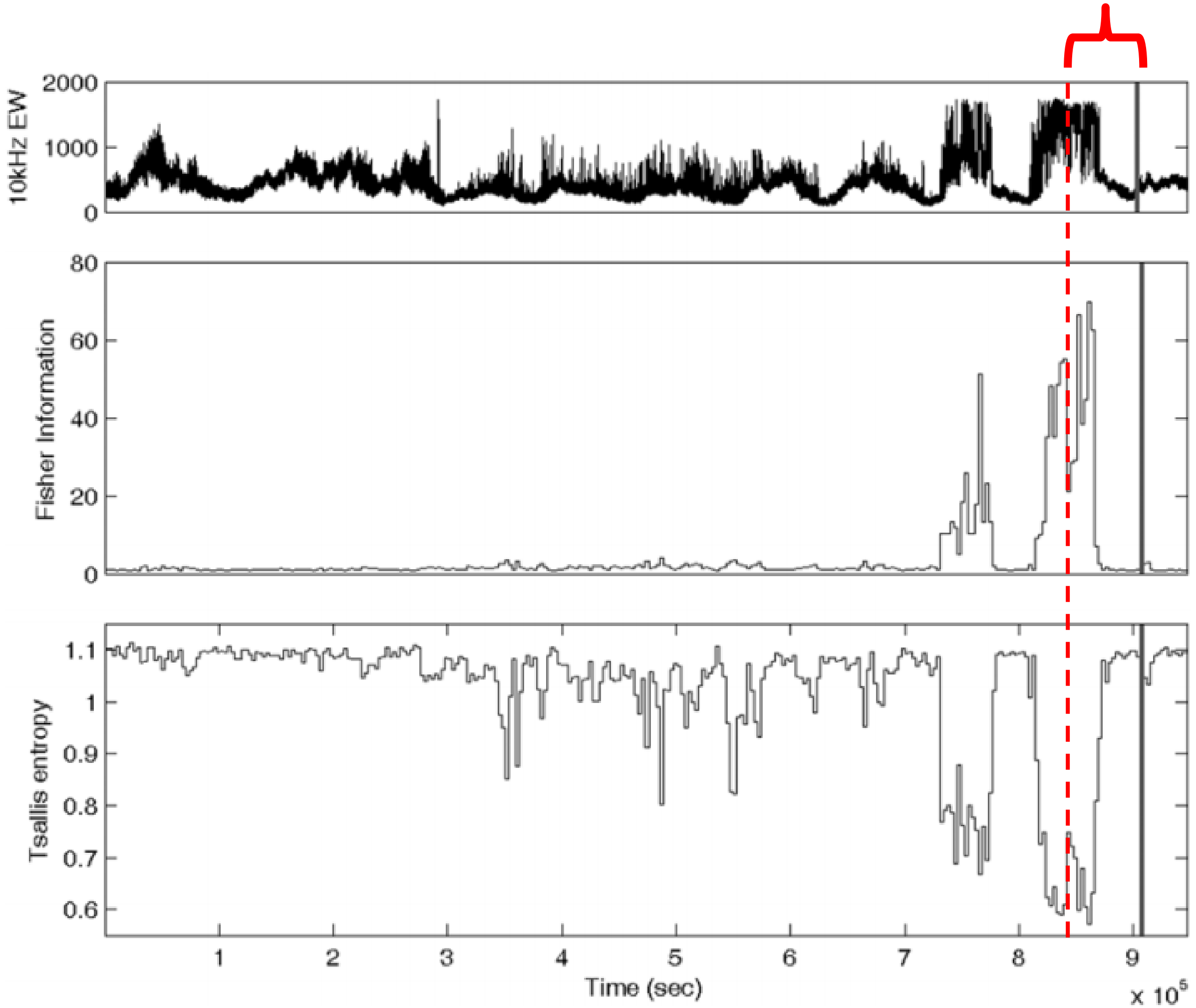
$$R_q[P] = \frac{1}{1-q} \ln \left[\sum_{j=1}^{n} (P_j)^q \right]$$

Fisher

$$F[P] = F_0 \sum_{j=1}^{n} \left[(p_{j+1})^{1/2} - (p_j)^{1/2} \right]^2$$

Mediciones del campo magnético terrestre antes de un sismo.

13 horas antes



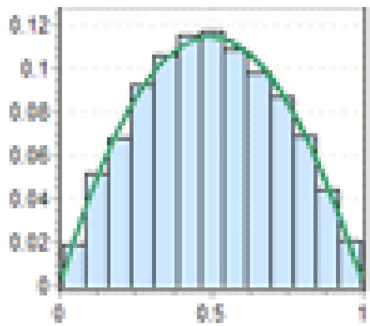
S.M. Potirakis Physica A (2012)

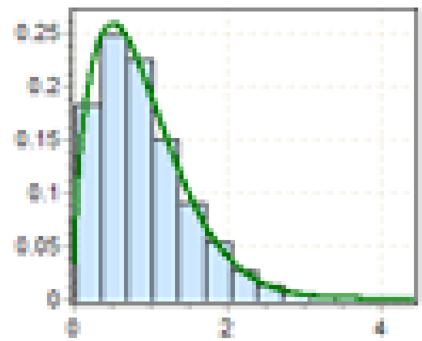
Divergencias

Divergencias estadísticas

Es la distancia relativa entre 2 distribuciones de probabilidad

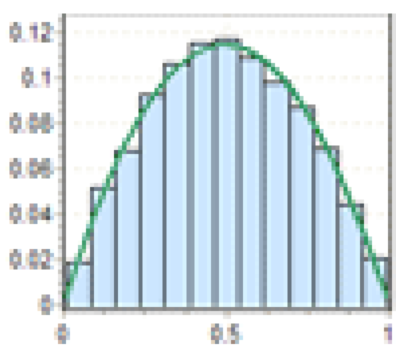
$P(x)$ $Q(y)$

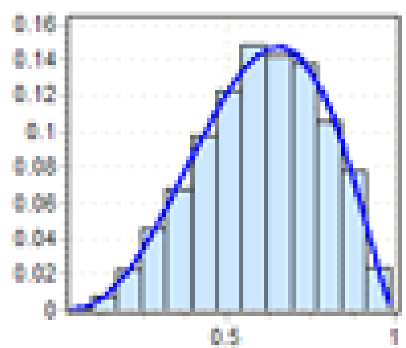




$D(P,Q) \rightarrow Alta$

$P(x)$ $R(y)$





$D(P,Q) \rightarrow 0$

Divergencia Kullback-Leible

Si tenemos dos variable aleatorias $X \, e \, Y$ las cuales puede tomar los valores $\{x_1, \; x_2, \ldots, x_n\} \, e \, \{y_1, y_2, \ldots, y_n\}$ con distribuciones de probabilidad asociada $P(X)$, $P(x_i) = P_i$ y Q(Y), $Q(y_i) = Q_i$ se define la divergencia Kullback-Leible (o entropía conjunta) como :

$$D_{KL}(P,Q) = \sum_{i=1}^{n} P_i \log\left(\frac{P_i}{Q_i}\right)$$

Si simetrizamos la D_{KL} se obtiene:

Divergencia: Jensen-Shannon

$$D_{JS}(P,Q) = H\left(\pi_1 P + \pi_2 Q\right) - \pi_1 H\left(P\right) - \pi_2 H(Q)$$

Se puede extender a n distribuciones de probabilidad

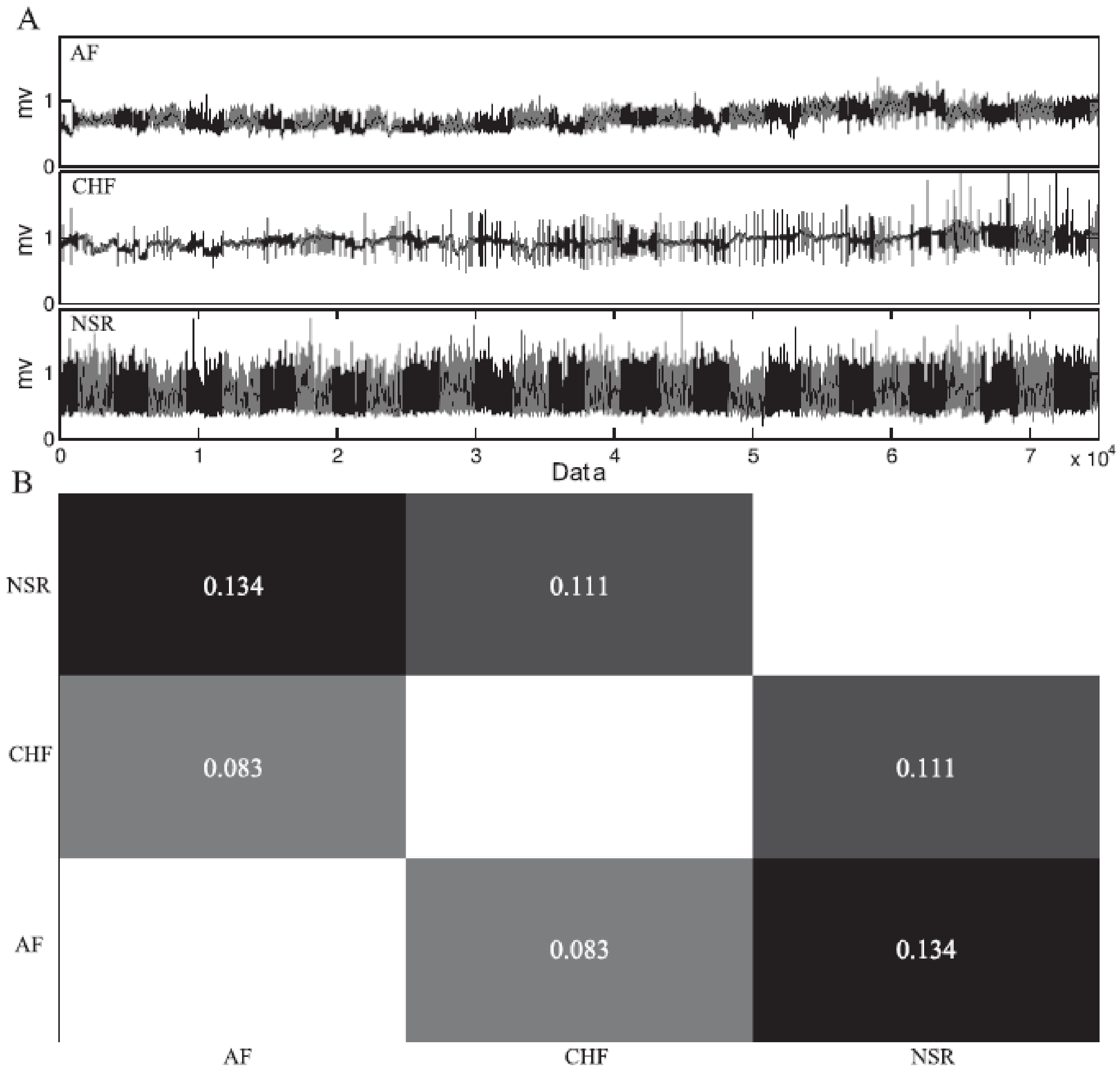
$$D_{JS}(P_1, P_2, \ldots, P_n) = H\left(\pi_i \sum_{i=1}^{n} P_i\right) - \sum_{i=1}^{n} \pi_i H\left(P_i\right)$$

Cumpliendo que : $\sum_{i=1}^{n} \pi_i = 1$

Para el $\log_2(\quad)$ *posee una cota :*

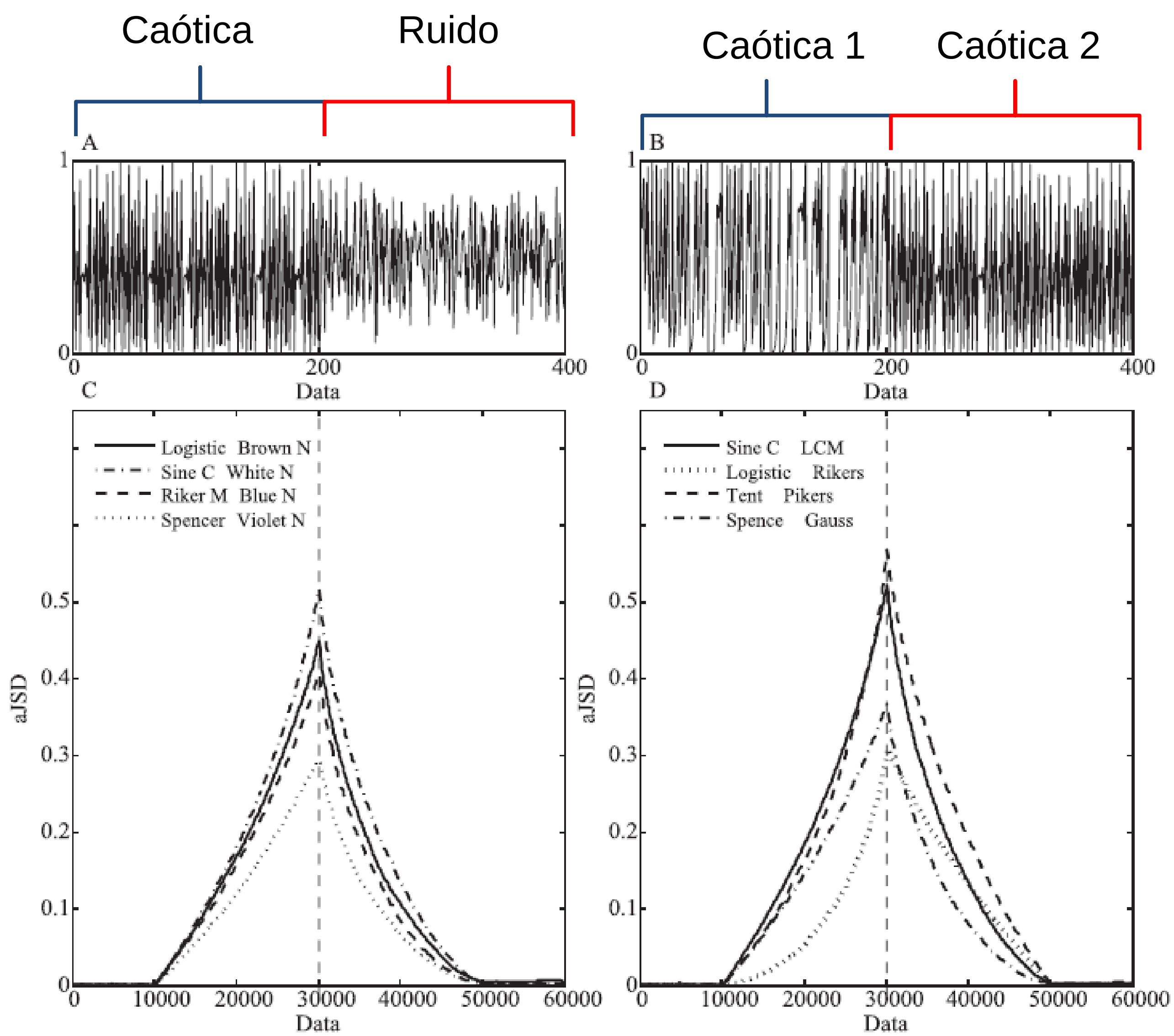
$$0 \le D_{JS}(P_1, P_2, \ldots, P_n) \le \log_2(n)$$

Distancia entre distribuciones de probabilidad de señales de ECG



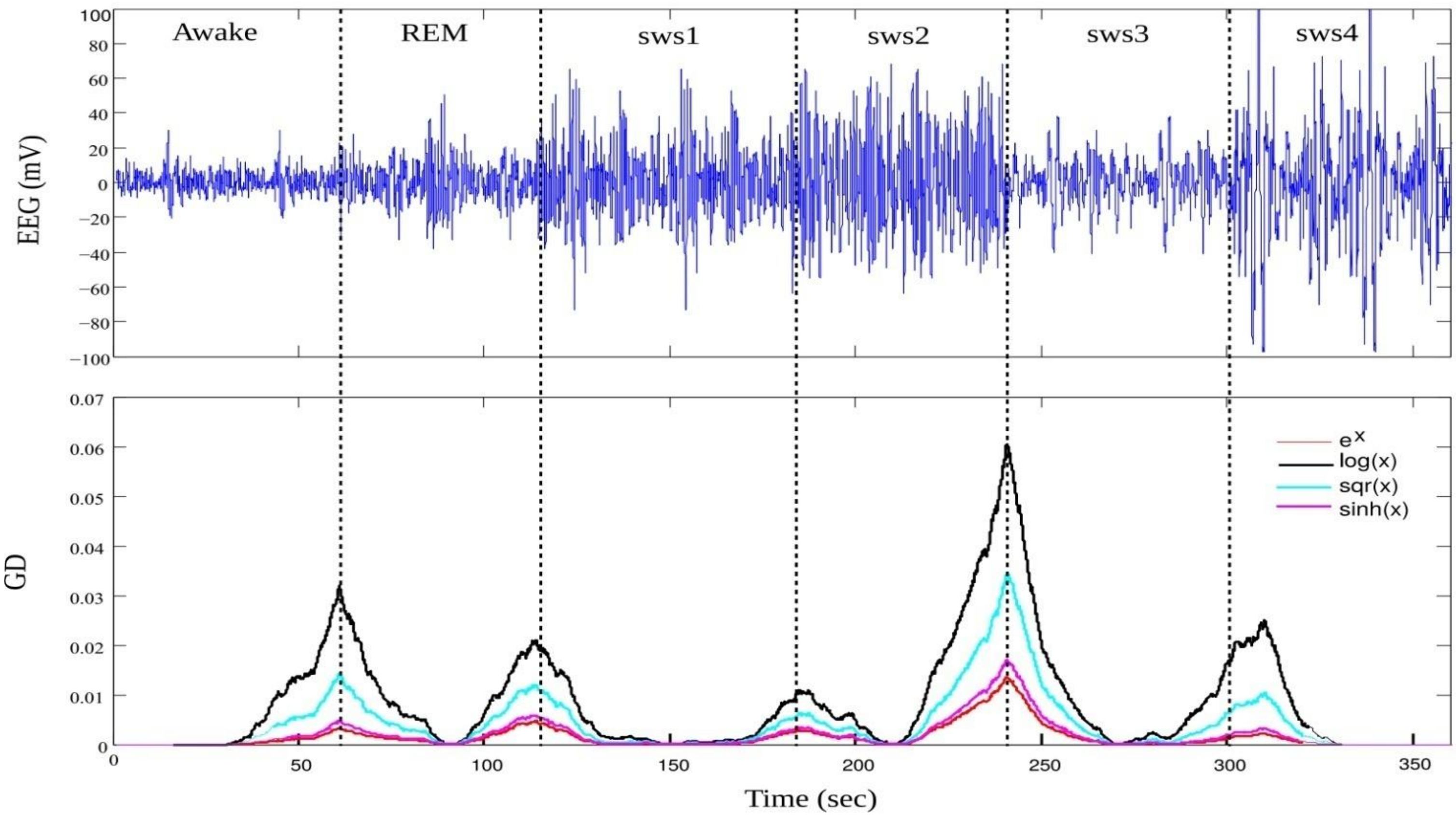
Mateos et. al. Chaos 2017

Distinción entre señales caóticas y Ruidos



Mateos et. al. Chaos 2017

Detección de los estadios de sueño en señales de polisomnografía



Riveaud et.al. Sin publicar 2018

Complejidades

Complejidad de una señal

↓

Es muy difícil tener una definición única de complejidad, ente otras cosas por que depende.

- Del sistema al que se aplica.
- De las herramientas matemáticas que se utilizan.
- Con que se lo compara.

Estadística

Algorítmica

Complejidad Estadística

Dada la variable aleatorea X con una distribución de probabilidad asociada $P(X)$, $P(x_i) = P_i$ y dada la distribución de probabilidad uniforme P_u se define la Complejidad como:

$$C_E(X) = Q_0 . D_{JS}(P, P_u) . H_n(P)$$

$$Q_0 = -2\left\{\frac{N+1}{N} . \ln(N+1) - 2\ln(2N) + \ln(N)\right\}$$

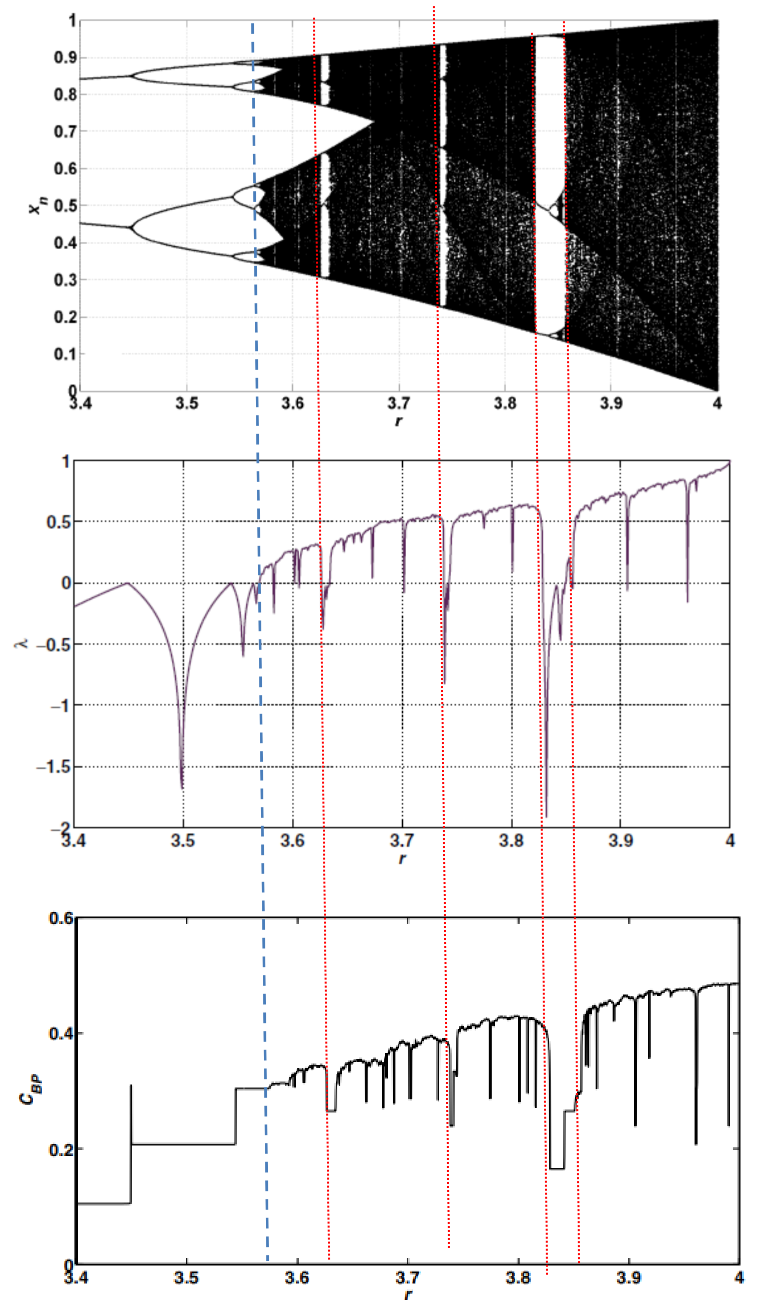
$$H_n = \frac{H(P)}{\ln(N)}$$

N = nº de estados posible del sistema

Caracterización Mapa Logístico

$$X_{n+1} = r.X_n(1 - X_n)$$

- La C se incrementa al aumentar la caoticidad.
- Es máxima para r=4 donde esta totalmente desarrollado el chaos.
- Los valores de C caen cuando están asociados a la ventanas periódicas



Complejidad de Lempel-Ziv

''Es una complejidad algorítmica que cuantifica la información no redundante en una secuencia''

- Se deriva de la noción de complejidad de ***Kolmogorov***

- La idea es reproducir una señal utilizando la menor cantidad de información , mediante la **producción y reproducción** de secuencias.

Lempel and Ziv. Transactions on information theory, 1978

Supongamos que tenemos la secuencia $S = 000110100$

0 → Símbolo nuevo, *producción*

0.0

0.00

0.001 → Símbolo nuevo, *producción*

$0.001.1$

$0.001.10$ → Símbolo nuevo, *producción*

$0.001.10.1$

$0.001.10.10$

$0.001.10.100$ → Símbolo nuevo, *producción*

Historia $(Hist)$

El numero de producciones es la complejidad de la historia $C_{hist}(S) = 4$

La complejidad de Lempel-Ziv se define como:

$$C_{LZ}(S) = \min\{C_{hist}(S)\}$$

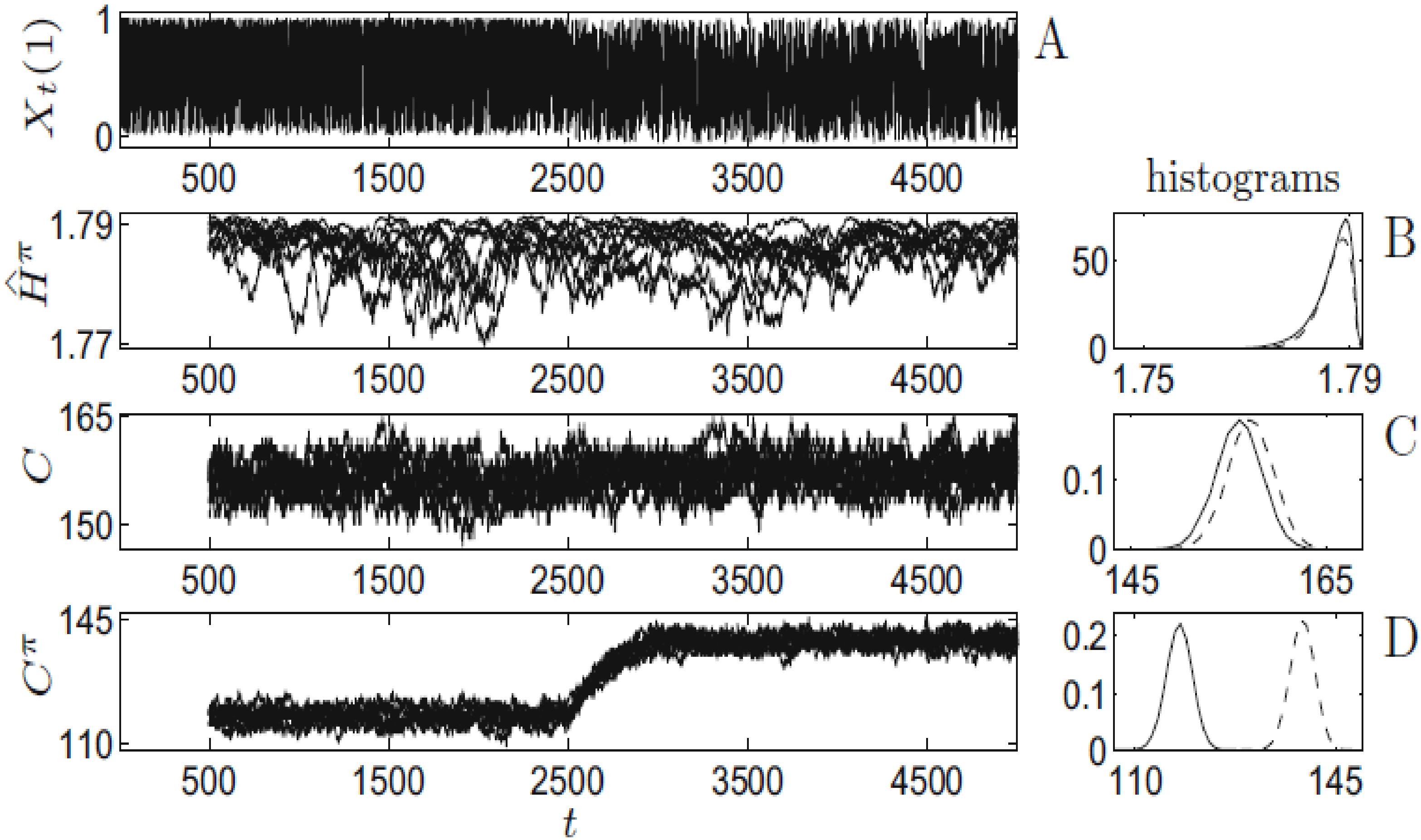
* Una propiedad importante es que se puede utilizar en **secuencias cortas**. Algo que no se puede hacer con las medidas de base estadística

Complejidad de Lempel Ziv y Entropia de Shannon

Cuando se tiene un proceso aleatorio **estacionario** y **ergódico**, normalizando correctamente, la complejidad de Lempel-Ziv de la secuencia tiende a la tasa de entropía de la misma:

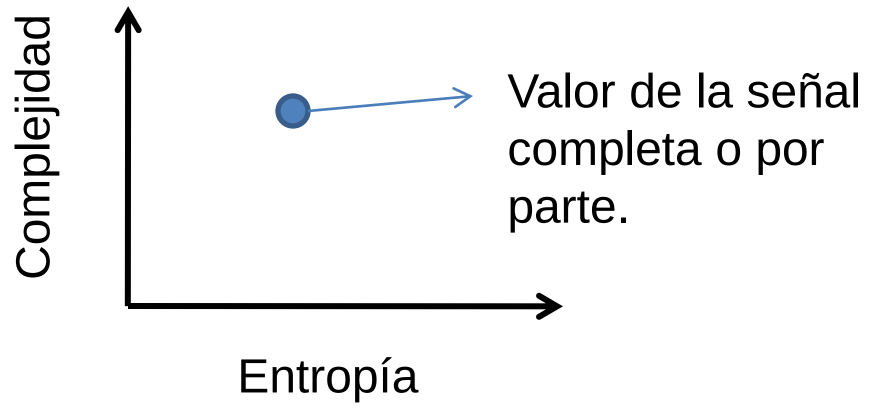
$$\lim_{T \to +\infty} C_{LZ}(S) \frac{\log(T)}{T} = \lim_{T \to +\infty} \frac{H(S)}{T}$$

Distinción entre una señal logística 3D y un ruido Blanco



Plano complejidad-entropía

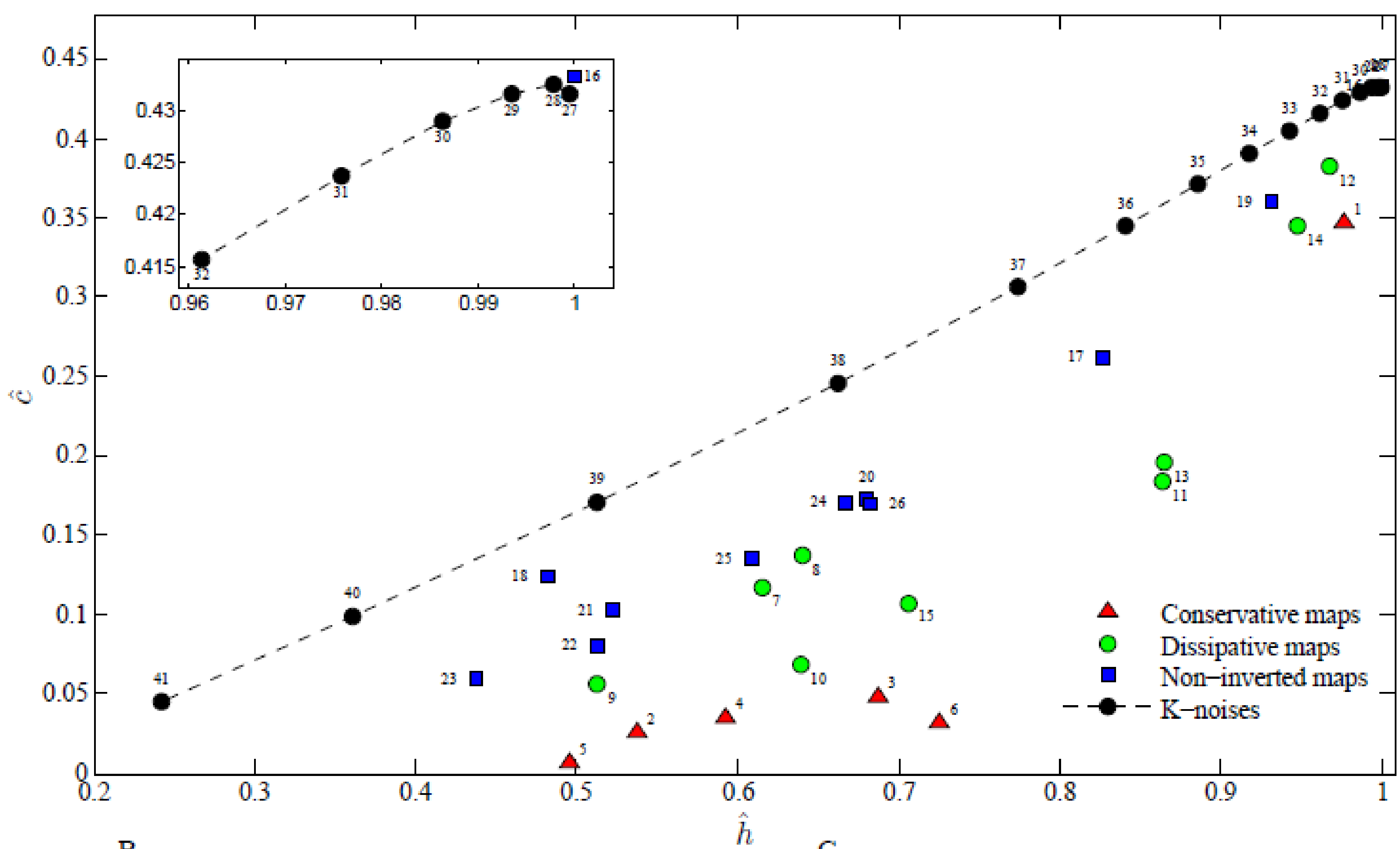
Que es un plano complejidad entropía?



Que ventajas tiene ?

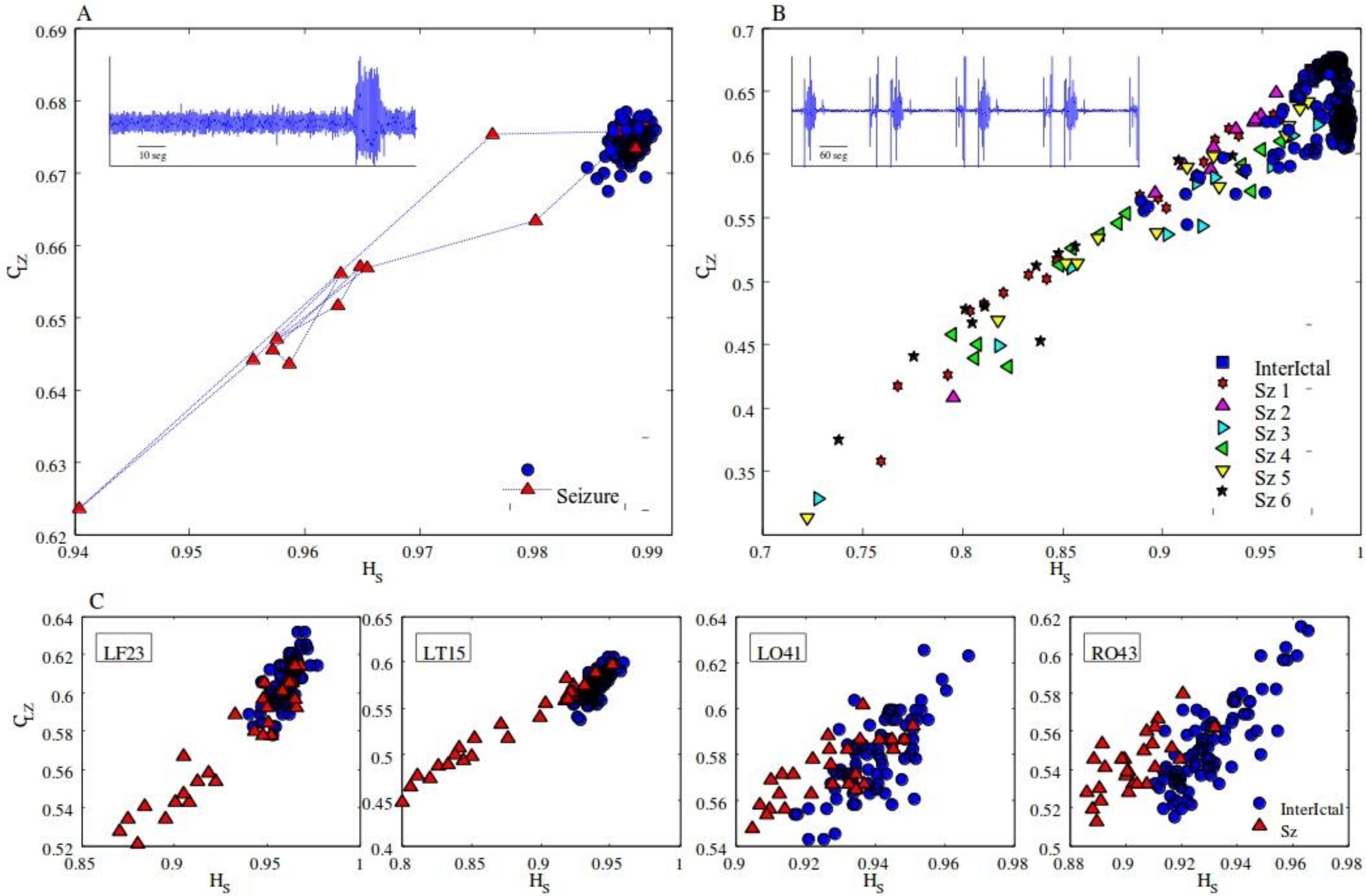
- Poder extraer mayor cantidad de información de las señales
- Mejor visualización de los datos.
- Poder distinguir diferentes tipos de señales según su dinámica.
- Mejor comprensión de los resultados, de una manera fácil y sencilla

Caracterización de mapas caóticos

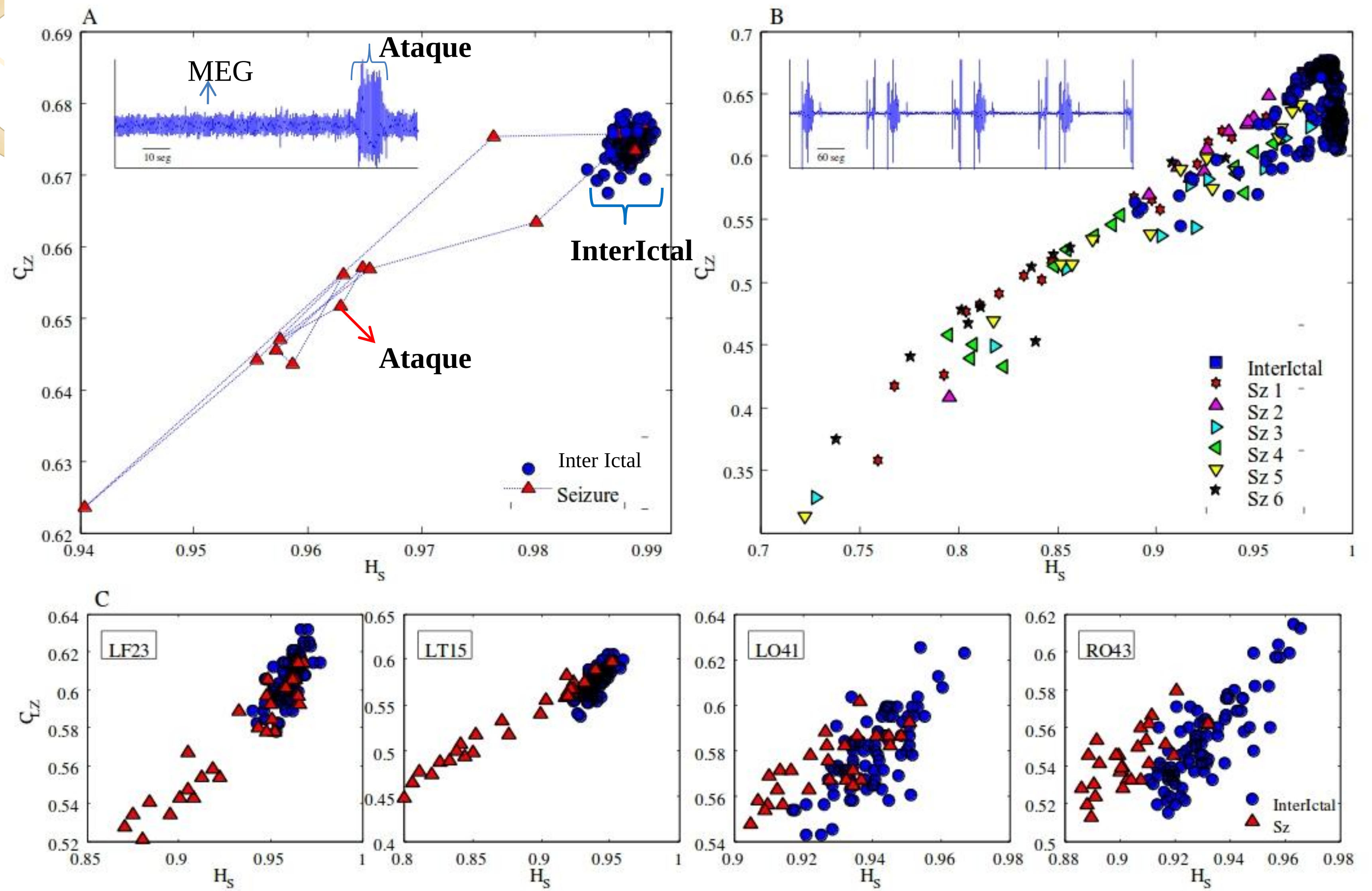


B C

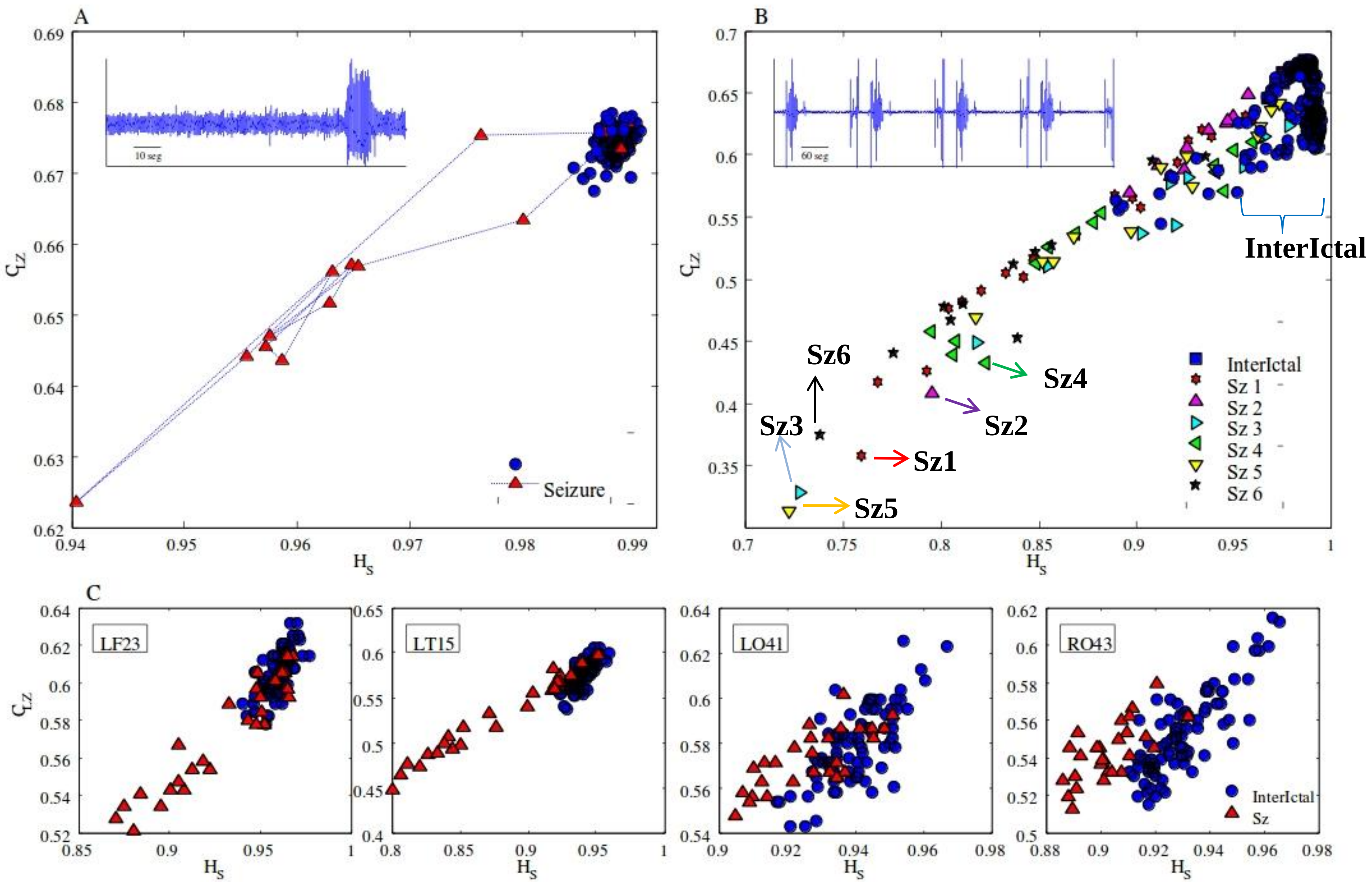
Análisis de la dinámica de las señales cerebrales en ataques epilépticos



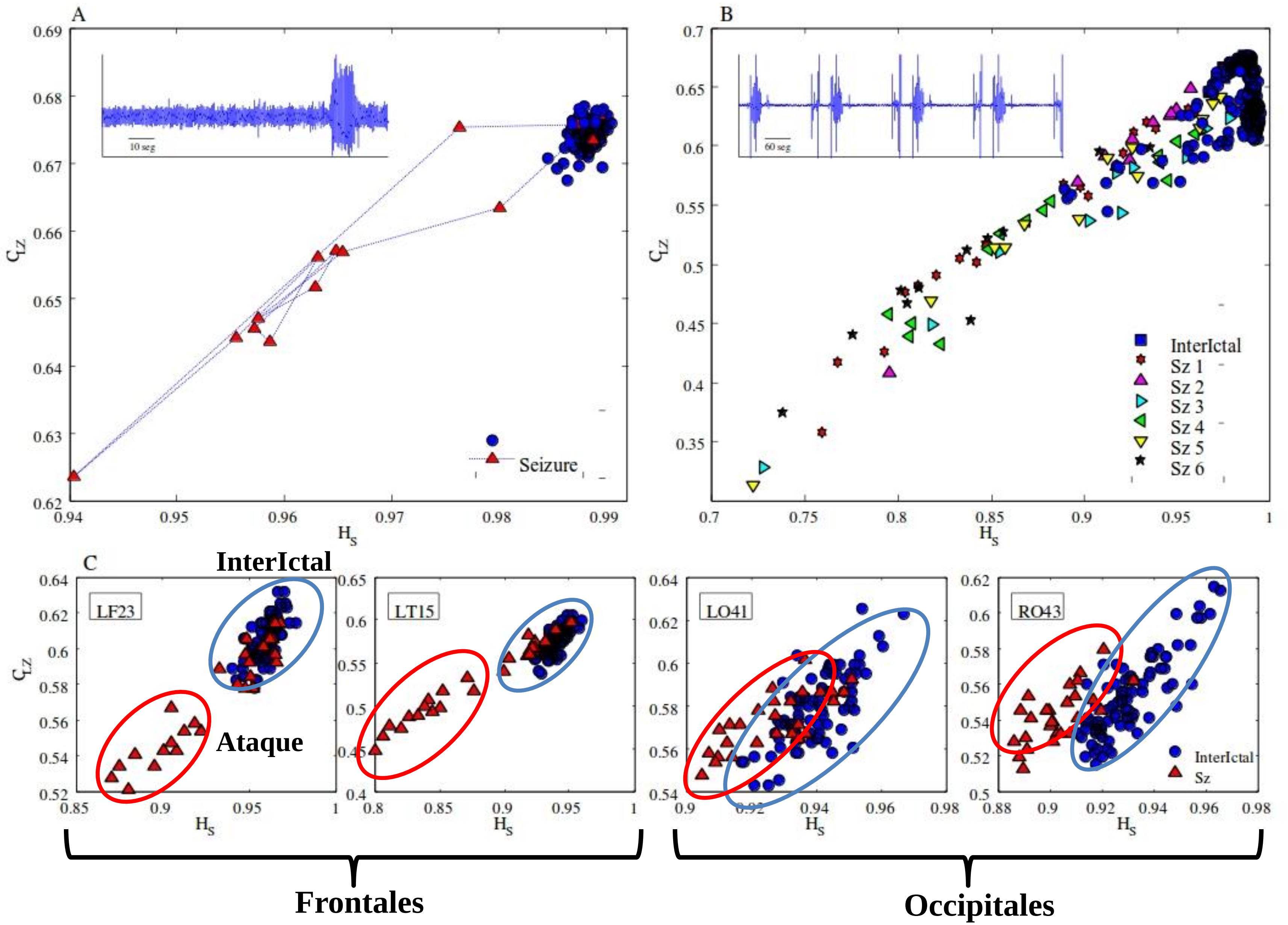
Análisis de la dinámica de las señales cerebrales en ataques epilépticos



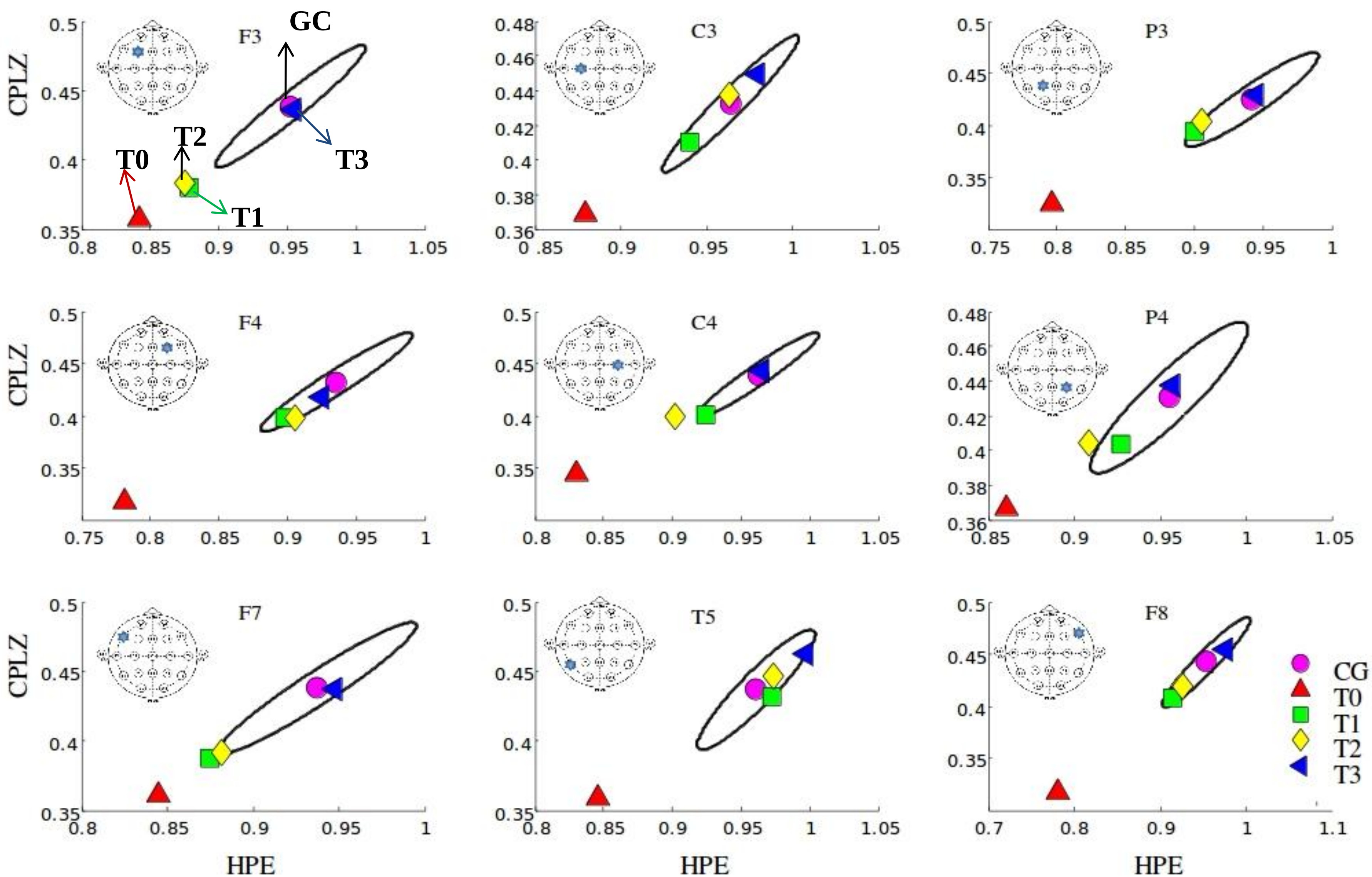
Análisis de la dinámica de las señales cerebrales en ataques epilépticos



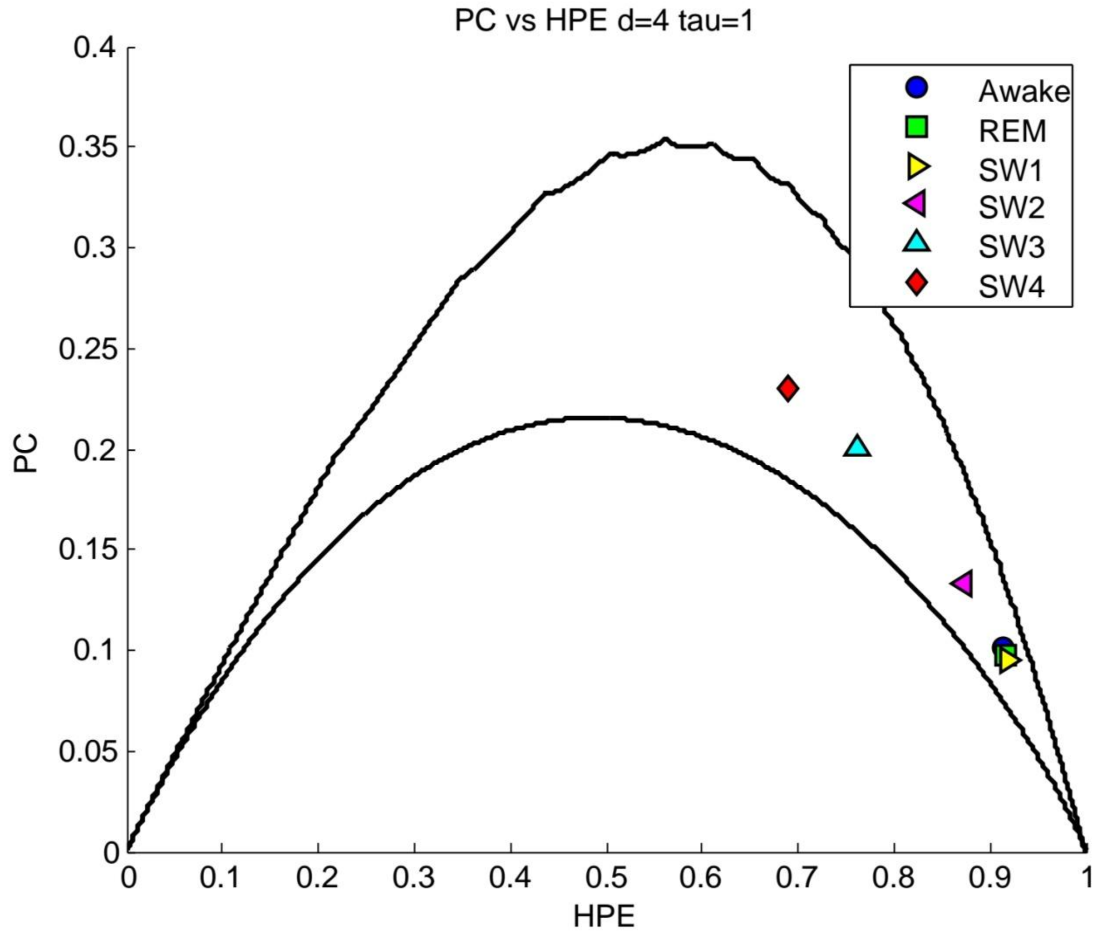
Análisis de la dinámica de las señales cerebrales en ataques epilépticos



Caracterización de la evolución clínica de pacientes bajo tratamiento framacológico.



Caracterización de estados de sueño.



Muchas gracias

¿Preguntas?