



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

---

FACULTAD DE CIENCIAS  
REDUCCIÓN DIMENSIONAL Y  
DETECCIÓN DE EVENTOS EN SERIES  
DE TIEMPO MULTIVARIADAS NO  
ESTACIONARIAS

TESIS

QUE PARA OBTENER EL TÍTULO DE:

**Físico**

PRESENTA:

**Noel Isaías Plascencia Díaz**

DIRECTOR DEL TRABAJO:

Marco Arieli Herrera Valdez

Ciudad Universitaria, Ciudad de México, 2019





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

# Índice general

|  |           |
|--|-----------|
| <b>1. Introducción</b>   | <b>4</b>  |
| 1.1. ¿Qué es la epilepsia? . . . . .   | 5         |
| 1.1.1. Descargas epilépticas interictales . . . . .  | 6         |
| 1.1.2. Causas comunes de la epilepsia . . . . .  | 6         |
| 1.2. Mecanismos de la epilepsia . . . . .  | 6         |
| 1.2.1. Actividad neuronal excesiva . . . . .   | 6         |
| 1.3. Diagnóstico y tratamiento de la epilepsia . . . . .                                   | 8         |
| 1.3.1. Pruebas utilizadas para diagnosticar epilepsia . . . . .                            | 9         |
| 1.4. Incidencia de la epilepsia . . . . .  | 10        |
| 1.5. Contexto histórico y moderno en el estudio de la epilepsia . . . . .                  | 11        |
| 1.6. Análisis de series de tiempo aplicado al EEG . . . . .                                | 13        |
| 1.6.1. ¿Qué es y cómo funciona un EEG? . . . . .   | 13        |
| 1.6.2. Naturaleza de las señales electroencefalográficas . . . . .                         | 13        |
| 1.6.3. Análisis computacional de series de tiempo . . . . .                                | 14        |
| 1.6.4. Series de tiempo multivaluadas . . . . .  | 15        |
| 1.6.5. EEG como una serie de tiempo multivaluada para el estudio de la epilepsia . . . . . | 15        |
| <b>2. Análisis de componentes principales</b>  | <b>20</b> |
| 2.1. Descripción analítica de PCA . . . . .  | 21        |
| 2.1.1. Criterios de reducción de dimensionalidad en PCA . . . . .                          | 24        |
| 2.2. PCA dependiente del tiempo . . . . .  | 25        |
| 2.2.1. PCA Dinámico . . . . .  | 26        |
| 2.2.2. PCA Recursivo . . . . .   | 27        |
| 2.2.3. PCA con ventana movible . . . . .   | 27        |
| 2.3. Aplicaciones de PCA al estudio de EEGs . . . . .                                      | 27        |
| <b>3. Resultados y discusión</b>   | <b>29</b> |
| 3.1. Obtención y descripción de los datos . . . . .  | 29        |

|  |           |
|--|-----------|
| 3.2. Aplicación de PCA en ventanas de tiempo . . . . .         | 31        |
| 3.3. Descripción del algoritmo . . . . .                       | 31        |
| 3.4. Descripción del análisis . . . . .                        | 34        |
| 3.4.1. Análisis de la varianza explicada . . . . .             | 34        |
| 3.5. Análisis de fluctuaciones de $V_E$ . . . . .              | 37        |
| 3.6. Clasificador automático de eventos . . . . .              | 43        |
| <b>4. Conclusiones</b>   | <b>47</b> |
| 4.1. Posibles extensiones . . . . .                            | 50        |
| 4.1.1. Reconstrucción del espacio fase en series de tiempo . . | 50        |
| 4.2. Agradecimientos . . . . .                                 | 53        |

# Capítulo 1

## Introducción

La epilepsia destaca como un fenómeno que persiste sin ser entendido por completo; aunque haya sido observada desde hace mucho tiempo. El desarrollo científico en la medicina, la biología molecular y las neurociencias ha impulsado la incorporación de nuevos métodos y tecnologías para la investigación de la epilepsia. Sin embargo, los mecanismos específicos de este trastorno y sus manifestaciones a diferentes escalas de organización siguen sin ser entendidos por completo.

Desde el punto de vista de los sistemas dinámicos, Strogatz (2018), clasifica a la epilepsia dentro de los temas de frontera en el conocimiento, debido a la alta dimensionalidad y no linealidad subyacentes a la dinámica de este trastorno.

Actualmente, la proliferación en la ciencia de marcos de trabajo con una perspectiva interdisciplinaria ha permitido que el estudio de la epilepsia incorpore elementos provenientes de la biología, la medicina y el modelado estadístico de sistemas dinámicos no lineales para el estudio de esta enfermedad.

En este trabajo de tesis, se propone una metodología de análisis que complementa conocimientos de ciencias de la vida, clínica, electrofisiología, sistemas dinámicos no lineales y análisis computacional de series de tiempo para ayudar al médico detectar patrones de actividad epileptiforme en registros de pacientes humanos. La metodología aquí presentada puede ser aplicada al estudio de otros sistemas con un número grande de variables que presenten una alta complejidad espacio-temporal en su dinámica.

Los resultados de esta tesis sugieren que ciertos patrones de cambio asociados a la variación conjunta de las señales del EEG podrían ser buenos indicadores de actividad anormal asociada a epilepsia. Sin embargo, debido a la limitada cantidad de registros con los que se contó para realizar el análisis, los resultados aquí presentados están pendientes de confirmación.

## 1.1. ¿Qué es la epilepsia?

En tiempos antiguos la epilepsia era conocida como la *enfermedad de los dioses*, posiblemente debido a los característicos ataques repetidos que pueden hacer que el paciente epiléptico pierda la conciencia y el control de su cuerpo por un tiempo determinado antes de regresar a un aparente estado de normalidad. En la medicina moderna, la epilepsia es descrita como un conjunto de trastornos, caracterizados por la predisposición permanente a generar crisis epilépticas y por las consecuencias neurobiológicas, cognitivas y sociales propias de la enfermedad (Fisher et al., 2014). Las manifestaciones fisiológicas más conocidas de la actividad epiléptica corresponden a las denominadas *crisis* o *ataques epilépticos* en los cuales un sujeto afectado puede perder el control de los músculos, así como sufrir pérdida de la conciencia.

Aunque los mecanismos fisiológicos subyacentes a esta enfermedad no se conocen con exactitud (Jasper, 2012), se sabe que los ataques epilépticos se originan a partir de la depolarización *en cadena* de un gran número de neuronas. A su vez, esto se debe a un aumento en la probabilidad de depolarización neuronal a lo largo de los circuitos neuronales de forma que se puede afectar a diferentes áreas del cerebro. Existen diversos tipos de ataques epilépticos y estos dependen en gran parte de la zona del cerebro en la que está localizado el o los focos epilépticos (Kandel et al., 2000).

Debido a que el conocimiento actual es insuficiente para clasificar las crisis epilépticas en términos de mecanismos fisiológicos específicos, se ha adoptado una clasificación operacional (Fisher et al., 2014) que toma en cuenta el lugar del cerebro en el que comenzó la crisis, alteraciones observables en el estado de alerta durante la crisis, y otras características como la presencia de movimientos musculares involuntarios.

Existen factores que pueden facilitar la presencia de crisis, entre estos se encuentran el consumo de alcohol, el estrés, la falta de sueño reparador, y la exposición a ciertos estímulos visuales.

### 1.1.1. Descargas epilépticas interictales

Además de las manifestaciones epilépticas en forma de crisis, se han caracterizado patrones de actividad denominados *descargas epilépticas interictales* (DEI), que son tipos de actividad que suelen detectarse en el 95 % de los pacientes con epilepsia (Asensio (2010)) y que pueden preceder a la aparición de una crisis epiléptica clínica.

### 1.1.2. Causas comunes de la epilepsia

Dentro de las causas más comunes de esta enfermedad se encuentran los traumatismos craneoencefálicos, infecciones en el sistema nervioso central, tumores cerebrales y factores genéticos, entre otras (WebMD, 2019). Sin embargo, debido a la naturaleza multifactorial de la epilepsia, en ocasiones no se logra identificar cuales son las causas específicas que provocan el desarrollo de la epilepsia en algunos pacientes.

## 1.2. Mecanismos de la epilepsia

Para estudiar y entender las diferentes dinámicas de actividad cerebral que caracterizan a la epilepsia es necesario tener en cuenta que las neuronas están organizadas en grupos que pueden tener millones de ellas, cada uno con funciones específicas y que a su vez están interconectados (Pessoa (2014)). En lo que queda del texto nos referiremos de forma indistinta a estos conjuntos de neuronas como *grupos, circuitos o redes* neuronales.

Mucha de la actividad cerebral normal ocurre de forma que los circuitos neuronales excitan su actividad mutuamente y en forma secuencial. En contraste, la actividad epileptiforme se caracteriza porque un gran número de neuronas actúan disparando de forma excesiva en un mismo intervalo de tiempo (Jefferys, 2017); (Hammer and McPhee, 2014). A los circuitos neuronales en donde se inicia la actividad epiléptica se les denomina *focos epilépticos*.

### 1.2.1. Actividad neuronal excesiva

Un tipo de actividad neuronal que caracteriza a la epilepsia se debe a la descarga eléctrica (o depolarización) en forma de ráfaga (Jefferys, 2017)

de un gran número de neuronas en un mismo intervalo de tiempo (Hammer and McPhee, 2014). Este fenómeno es conocido como cambio paroxístico de depolarización o *paroxysmal depolarizing shift* (PDS) (Fig. 1.1). Después de este episodio suele presentarse un periodo conocido como depresión post-ictal, que se caracteriza por una disminución generalizada en la actividad eléctrica neuronal.

Los patrones de actividad descritos anteriormente son conocidos de forma más general como descargas epilépticas interictales (DEI) debido a que suelen preceder a la aparición de una crisis epiléptica clínica. Se considera que las DEI, en un adecuado contexto clínico son marcadores electrofisiológicos de epilepsia, pues sólo alrededor del 2 % al 4 % de los EEGs de un individuo normal presentan DEIs (Asensio, 2010).

Los perfiles de actividad anormal característicos de la epilepsia pueden deberse a patologías en los canales iónicos de las células o al mal funcionamiento de neuronas inhibitorias. Dichos cambios fisiológicos pueden traducirse respectivamente en un aumento en la excitabilidad después de la depolarización de las neuronas y una disminución en la inhibición en las mismas. Esto puede provocar un aumento en la probabilidad de depolarización de las neuronas vecinas. También es posible encontrar una combinación de estos dos mecanismos como causantes de actividad epileptiforme (Jefferys, 2017).

En el caso de la epilepsia el concepto de hiperactividad juega un papel esencial. Una de las características de esta enfermedad es el aumento de la excitabilidad de un grupo de neuronas, es decir, un aumento en la probabilidad de disparo de estas y de neuronas vecinas, provocando que muchas neuronas disparen en un mismo intervalo de tiempo, en un efecto de retroalimentación positiva que suele ser comparado con una avalancha y el cual puede corresponder con un episodio clínico de crisis epiléptica.

Los fenómenos de hiperactividad en epilepsia son resultado de una combinación de mecanismos fisiológicos, destacando la acumulación de calcio causada por la depolarización de neuronas cercanas al foco epiléptico y cambios en los niveles de cloro en los canales neuronales, entre otros posibles mecanismos. Además efectos de campo eléctrico que también pueden contribuir en el aumento de la actividad neuronal (Hammer and McPhee, 2014); (Siegel et al., 1994).



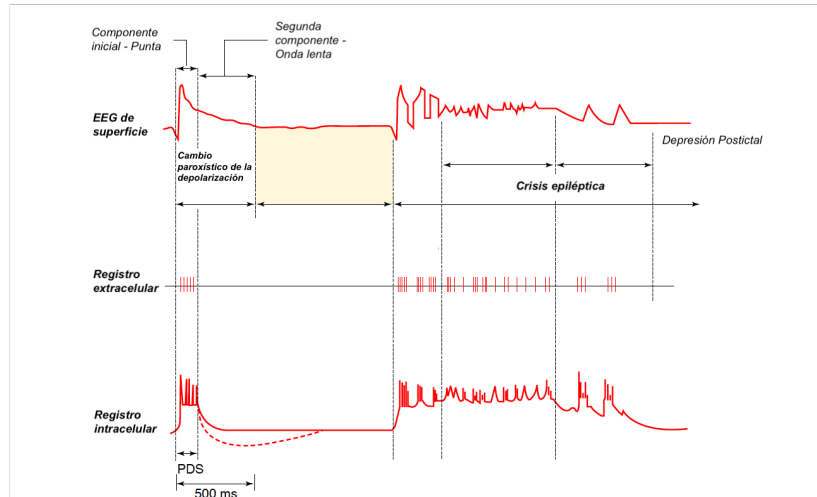


Figura 1.1: *Actividad epiléptica representada a diferentes niveles de organización. En las partes correspondientes a registros intra y extracelulares podemos observar que esta actividad se manifiesta como disparo en forma de ráfaga durante el PDS, seguida de un estado de inactividad de la neurona y, posteriormente, actividad de disparo relacionada con una crisis epiléptica (SEIZURE), que finaliza con la inactividad de la neurona (Postictal depression). En el registro de EEG (traza superior del esquema) el PDS se manifiesta como una punta de actividad -denominada descarga epiléptica interictal (DEI)-, seguida de un periodo refractario que a suele preceder a la crisis epiléptica clínica (Siegel et al. (1994)). Imagen adaptada de Hammer and McPhee (2014).*

### 1.3. Diagnóstico y tratamiento de la epilepsia

De acuerdo con definición establecida en 2005 por la Liga Internacional en Contra de la Epilepsia (ILAE), el diagnóstico de epilepsia se puede establecer en las siguientes condiciones (Fisher et al. (2014)):

1. La presencia de 2 crisis no provocadas mediante algún factor determinante (como un traumatismo o hipoglicemia, por ejemplo) separadas al menos por 24 horas.
2. Una convulsión no provocada más el riesgo de presentar un nuevo episodio epiléptico cuando menos de un 60% después de dos convulsiones no

provocadas dentro de los 10 años siguientes.

3. Diagnóstico de un síndrome epiléptico, es decir, la presencia de un conjunto de características epilépticas específicas, como la epilepsia de ausencias en la adolescencia o la epilepsia mioclónica juvenil.

Para emitir un diagnóstico correcto, el médico especialista tiene que descartar la posibilidad de que dichos ataques se deban a causas ajenas a la epilepsia como en el caso de las crisis de origen no epiléptico, algunos tipos de migraña, síncope y crisis psicógenas.

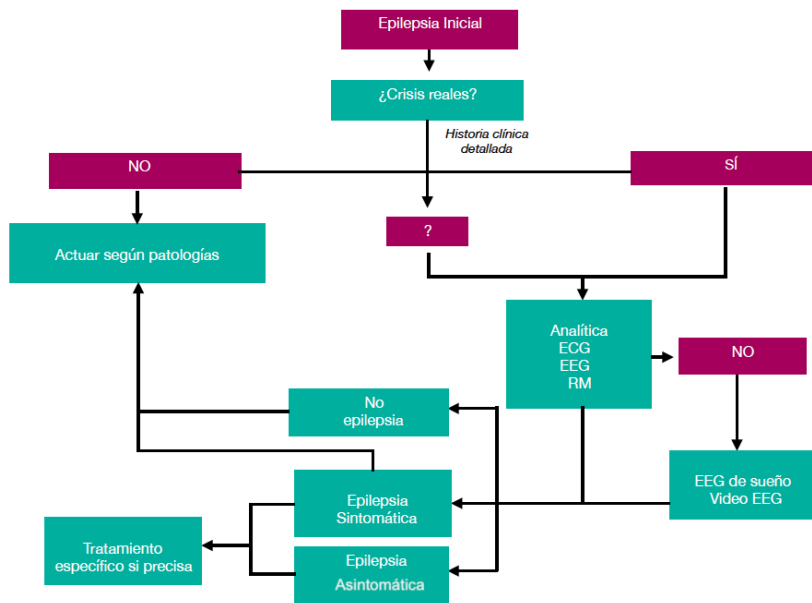


Figura 1.2: *Protocolo general para el diagnóstico epilepsia, imagen adaptada de Asensio (2010).*

### 1.3.1. Pruebas utilizadas para diagnosticar epilepsia

Para determinar qué es lo que está provocando las crisis en el paciente, el médico debe de conocer la historia clínica detallada del mismo, además de apoyarse en protocolos de diagnóstico especializados. Las pruebas más comunmente requeridas por los médicos son (Holland, 2015):

En primer lugar, el conocimiento del momento y contexto en el que iniciaron las crisis en el paciente. Esto puede ayudar al médico a elegir las pruebas de diagnóstico adecuadas, así como un posible tratamiento.

Las *pruebas de sangre* son requeridas para detectar posibles infecciones que puedan explicar los síntomas epilépticos, una infección en el sistema nervioso puede ser causante de la epilepsia.

El *Electroencefalograma (EEG)* es la herramienta de diagnóstico de epilepsia más eficiente, este consiste en colocar electrodos en el cuero cabelludo o directamente en la corteza cerebral del paciente para registrar la actividad eléctrica de grandes grupos de neuronas. Este registro es analizado posteriormente por un especialista para encontrar actividad inusual, la cual podría estar ligada a la presencia de epilepsia.

La *tomografía computarizada* es una técnica de imagenología que puede ayudar al especialista médico a encontrar posibles causas de crisis epilépticas como tumores, quistes, aneurismas e indicios de eventos vasculares cerebrales.

La *imagen por resonancia magnética (IMR)* es otra técnica de imagenología usada para el diagnóstico, el médico busca las mismas causas de epilepsia que en una tomografía computarizada con la diferencia de que la IMR ofrece mayor precisión.

De ser confirmada la presencia de un trastorno epiléptico, el médico indica el tratamiento farmacológico, este es personalizado para cada paciente, ya que depende del tipo de crisis, edad, comorbilidades y estilo de vida del mismo (NICE, 2017). En los países subdesarrollados el 80% de los pacientes no reciben tratamiento o reciben tratamiento inadecuado. Latinoamérica consume el 7% de los fármacos antiepilépticos del mundo.

## 1.4. Incidencia de la epilepsia

La epilepsia es una de las enfermedades neurodegenerativas más recurrentes. De acuerdo con datos de la Organización Mundial de la Salud (OMS), esta enfermedad afecta a alrededor de 50 millones de personas en el mundo, esto es aproximadamente 1% de la población mundial.

## 1.5. CONTEXTO HISTÓRICO Y MODERNO EN EL ESTUDIO DE LA EPILEPSIA 11

Se estima que en la población general, existe una proporción de entre 4 y 10 pacientes con epilepsia por cada 1000 personas, cifra que aumenta de forma significativamente en países con ingresos medios y bajos, llegando a entre 7 y 14 por cada 1000 personas. Esto último podría estar relacionado con la prevalencia en estos países de enfermedades infecciosas que afectan al sistema nervioso como la malaria y la neurocisticercosis, así como la mayor incidencia de accidentes de tráfico o lesiones en el parto. Estudios realizados por la OMS indican que cerca del 80 % de las personas con epilepsia viven en países con bajo o medio desarrollo.

La OMS estima que globalmente, alrededor de 2.4 millones de personas son diagnosticadas por epilepsia cada año (WHO, 2017).

Estudios indican que la incidencia de epilepsia es mayor en niños menores de 10 años (debido a epilepsias sintomáticas, accidentes antenatales y neonatales, errores innatos del metabolismo, etc.) y en adultos mayores a 55 años (debido a eventos vasculares cerebrales y traumatismos craneoencefálicos) (Holland (2015)).

Se estima que la prevalencia de epilepsia en México es de 11 a 15 por cada 1000 personas, se calcula que el número de pacientes con epilepsia en México es cercano a 1.5 millones (HMG (2017)).

### 1.5. Contexto histórico y moderno en el estudio de la epilepsia

*”Los hombres creen que la epilepsia es divina simplemente porque no la entienden. Pero si llamaran divino a todo lo que no entienden, realmente las cosas divinas no tendrían fin.”*

-Hipócrates

Se considera que el análisis moderno de la epilepsia comenzó en los años 1860's con los trabajos de John Hughlings Jackson (Kandel et al., 2000), quien comenzó a establecer formalmente las diferencias entre las crisis focales, que pueden incluir movimientos involuntarios de los brazos o las piernas, y las crisis generalizadas, en las que los pacientes pueden presentar movimientos de todo el cuerpo e invariablemente se pierde la conciencia. Avances posteriores en el estudio neurobiológico de la epilepsia incluyen tratamientos quirúrgicos por Victor Horsley, quien en 1886 realizó una cirugía para remo-

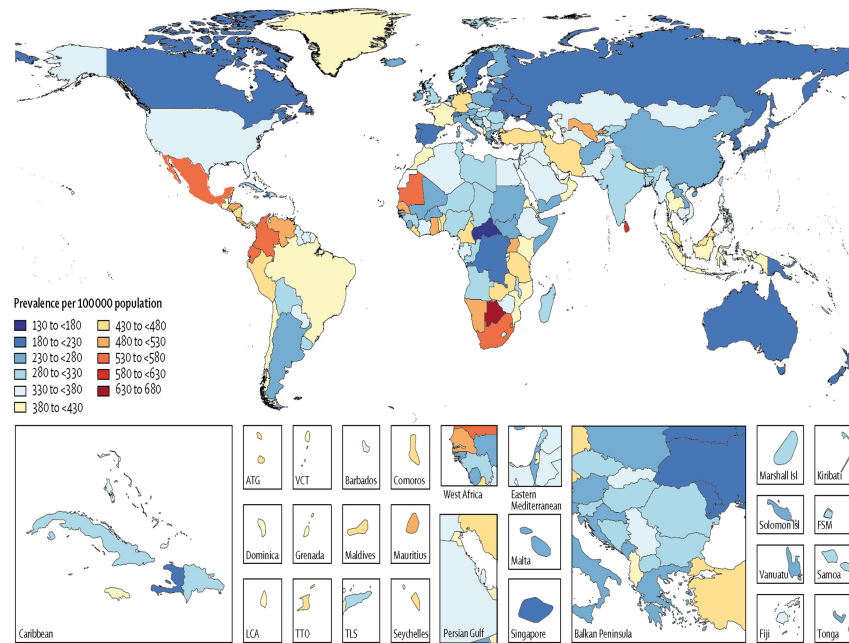


Figura 1.3: Prevalencia mundial de la epilepsia idiopática (sin causa identificable) por cada 100 000 personas para ambos sexos, 2016 (GBD (2016)).

ver tejido de la corteza cerebral adyacente a una fractura de cráneo, logrando curar a un paciente que presentaba epilepsia focal.

Los tratamientos clínicos modernos de la epilepsia comenzaron a surgir a inicios del siglo XX, estos incluyeron el uso de medicamentos para controlar las convulsiones (1912), el desarrollo de la electroencefalografía (1929) y el descubrimiento de la fenitoína, un fármaco que es usado como anticonvulsivo.

A causa de la naturaleza dinámica y altamente compleja de la epilepsia, el estudio de esta enfermedad ha permeado a otras disciplinas más allá de la medicina, como las matemáticas (Iasemidis and Sackellares, 1996), la física (Ebersole, 1991), la biología molecular (Lossin et al., 2002) y las ciencias de la computación (Lytton, 2008). A partir de los estudios interdisciplinarios se han logrado grandes avances en el estudio de esta enfermedad, se conoce más acerca de la etiología de la epilepsia, se han desarrollado tratamientos más efectivos y se ha trabajado en la creación de métodos de análisis que

permitan alertar la posible presencia de crisis epilépticas con cierto rango de precisión.

## 1.6. Análisis de series de tiempo aplicado al EEG

En esta sección se abordará más profundamente qué es un electroencefalograma (EEG), cómo funciona y cómo es usado clínicamente para diagnosticar epilepsia. Después mencionaré cuáles son los métodos computacionales más utilizados actualmente para el análisis de EEGs y terminaré describiendo el algoritmo llamado *Principal Component Analysis* (PCA), en el cuál se basa la metodología propuesta en esta tesis.

### 1.6.1. ¿Qué es y cómo funciona un EEG?

El EEG es una herramienta de apoyo al diagnóstico de enfermedades de origen neuronal que se ha convertido en la herramienta más utilizada para diagnosticar epilepsia alrededor del mundo. El EEG registra la actividad cerebral de neuronas por medio de electrodos que se colocan en la superficie del cuero cabelludo o directamente sobre la corteza cerebral. Estas señales eléctricas son amplificadas y mostradas gráficamente en una computadora. Para completar el diagnóstico basado en EEG es necesario que un experto realice una inspección ocular de la actividad eléctrica registrada e interprete esta en términos clínicos. Es importante mencionar que existen casos en los que la epilepsia es difícil de diagnosticar, pues alrededor del 10% de los pacientes epilépticos no exhiben actividad epiléptica que pueda ser detectada con un EEG (Smith, 2005).

### 1.6.2. Naturaleza de las señales electroencefalográficas

Un EEG requiere que los electrodos sean localizados en el cuero cabelludo o la corteza cerebral del paciente. De esta manera es posible captar pulsos eléctricos que corresponden a la actividad de grupos grandes de neuronas (Kandel et al., 2000). Es importante mencionar que el EEG tiene una baja resolución espacial, esto se debe a que un EEG sólo puede captar información de regiones grandes del cerebro, haciendo imposible detectar la actividad de circuitos neuronales específicos.

Desde un acercamiento cuantitativo, las señales de un EEG de varios canales pueden ser tratadas como una serie de tiempo multivaluada (Harper et al., 1974). Las siguientes contienen una descripción general de las características

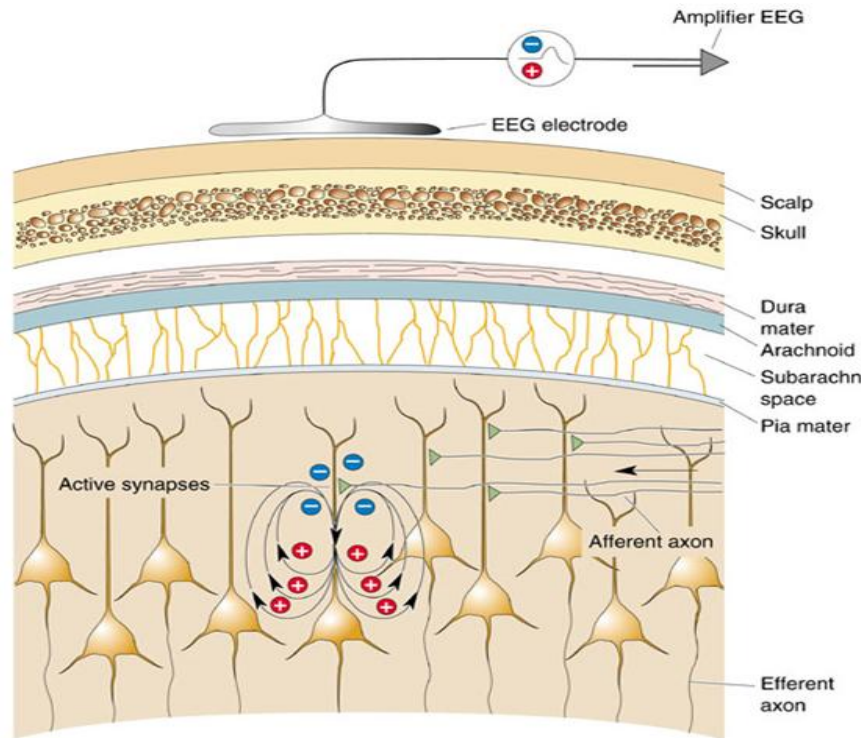


Figura 1.4: *Esquema representativo de la actividad eléctrica neuronal registrada por un EEG. El esquema muestra la depolarización de una neurona piramidal, provocando un impulso eléctrico que es registrado y amplificado por un electrodo colocado en la superficie del cuero cabelludo del paciente.* <https://121neurofeedback.com/wp-content/uploads/2017/07/brain-large.jpg>

de las series de tiempo y exponen la utilidad de estudiar al EEG como una serie de tiempo multivaluada.

### 1.6.3. Análisis computacional de series de tiempo

En matemáticas, una serie de tiempo de una variable de interés se define como un conjunto de observaciones cronológicamente ordenadas de dicha variable (Belsley et al., 1980). El ritmo cardíaco, los precios en la bolsa de valores, variables climatológicas como humedad y temperatura, la posición de un objeto y el número de personas que tienen una enfermedad son variables que fluctúan temporalmente y por lo tanto pueden considerarse como series de tiempo.

Las series de tiempo pueden contener cantidades muy grandes de datos (del orden de  $10^6$  o más por registro). Por esto mismo, se han desarrollado algoritmos computacionales para analizarlos de forma rápida y confiable. Estos algoritmos generalmente tienen como base la extracción y análisis de ciertas propiedades estadísticas de la serie.

#### 1.6.4. Series de tiempo multivaluadas

Una forma muy efectiva de estudiar series de tiempo es extraer características generales y/o patrones en los datos que permitan interpretar, clasificar o predecir el comportamiento de dicha serie. Sin embargo, este tipo de análisis puede volverse muy complicado cuando las series de tiempo son *multivaluadas*, es decir, cuando más de una magnitud observable varía en el tiempo. Un ejemplo de una serie de tiempo multivaluada de 3 dimensiones sería medir el ritmo cardiaco, la actividad muscular y la temperatura corporal de un corredor a lo largo del tiempo. Otro ejemplo son los electroencefalogramas (EEG), los cuales registran actividad cerebral en 21, 64 o 256 canales (dependiendo del equipo de registro utilizado). Tomando en cuenta la importancia de la variación conjunta de todas las señales, el hecho de tener una dimensión alta en la serie de tiempo puede ocasionar que extraer indicadores de interés de los datos sea una labor complicada mediante métodos de análisis tradicionales.

#### 1.6.5. EEG como una serie de tiempo multivaluada para el estudio de la epilepsia

El electroencefalograma (EEG) es una herramienta ampliamente usada en medicina como apoyo al diagnóstico, particularmente de trastornos epilépticos.

El análisis de electroencefalogramas se lleva a cabo a través de una exploración ocular del registro por parte de un médico experto. Sin embargo, este proceso puede llegar a tomar mucho tiempo, en especial en registros de larga duración, los cuales pueden extenderse desde unas cuantas horas hasta varios días. Finalmente, la naturaleza multifactorial de la epilepsia ocasiona que, a pesar de existir lineamientos para la interpretación clínica del EEG, este siga siendo un procedimiento subjetivo que depende en buena parte de la experiencia del médico. Lo anterior hace que la interpretación de los registros EEG sea propensa a errores humanos y, a su vez, ha motivado la



búsqueda de métodos de análisis más sistemáticos (Correa et al., 2015).



Figura 1.5: *Ejemplo de un registro de EEG de 22 canales. El EEG puede ser estudiado como una serie de tiempo multivariada con tantas dimensiones como canales de registro sean implementados en el montaje.*

La epilepsia es una enfermedad que involucra un gran número de variables en cada paciente (edad, sexo, estado de vigilia, presencia de comorbilidades, el tipo de medicamento usado, entre otras) y cuyos mecanismos fisiológicos subyacentes aún no se han entendido por completo. Además, la presencia o ausencia de patrones encefalográficos anormales en un registro no determina por sí misma el diagnóstico de epilepsia (Steven C. Schachter, 2014). A causa de esto, el médico experto, quien toma en cuenta la mayoría de estas variables, sigue siendo el mejor clasificador y analista de EEGs. Aun así, es posible implementar metodologías computacionales de análisis de EEG cuya finalidad sea el apoyo al diagnóstico por parte del médico experto.

Como se mencionó en la Sección 1.1.1, las interacciones entre circuitos neuronales involucran fenómenos que se regulan a lo largo de distintos niveles de organización, los cuales abarcan desde procesos a nivel molecular hasta interacciones eléctricas a nivel macroscópico. La no linealidad de los procesos fisiológicos provoca que las dinámicas de interacción entre circuitos neuronales y la relación de estos últimos con la presencia de patologías a nivel clínico resulte en un problema difícil de abordar.

Macroscópicamente, en el EEG, esto se ve reflejado en la no linealidad característica de estos registros. Además es sabido que las series de tiempo de EEG son no estacionarias, es decir, las propiedades estadísticas de ésta cambian con el tiempo. Además de estas características está la presencia de ruido en el EEG.

Las características anteriores de los registros EEG hacen que, para personas no entrenadas, la actividad eléctrica registrada en un EEG sea difícil de clasificar e interpretar. En el caso de un cerebro con epilepsia, este problema se traduce en la dificultad de detectar actividad epileptiforme en un EEG y establecer indicadores que puedan ayudar a predecir crisis epilépticas.

En resumen, las características que dificultan el estudio de señales electroencefalográficas son esencialmente 3 (Klonowski, 2009): **no-linealidad, no estacionariedad y presencia de ruido externo**. Hacer uso de técnicas, conceptos y herramientas analíticas que tomen en cuenta las tres características anteriores podría ser el acercamiento más adecuado para estudiar señales de EEG de forma cuantitativa (Stam (2005), Micheloyannis et al. (1998a)).

Uno de los objetivos principales de analizar EEGs computacionalmente es encontrar patrones de actividad eléctrica que indiquen la presencia de enfermedades de forma confiable y eficiente. Debido a que distintas enfermedades a nivel cerebral manifiestan distintos patrones de actividad en un EEG, es necesario encontrar algoritmos de análisis adecuado para cada caso considerando que la señal (o serie de tiempo) del EEG es altamente no lineal y no estacionaria (Stam et al. (1999)).

En la actualidad, se han desarrollado una gran cantidad de métodos y algoritmos para analizar computacionalmente registros de EEG con epilepsia, el rango de métodos utilizados comprende desde la aplicación de análisis lineales clásicos (como el análisis de Fourier) hasta el uso de herramientas y conceptos derivados de la teoría de sistema dinámicos no lineales, pasando por la implementación de modelos de aprendizaje de máquina y redes neuronales artificiales (Orosco et al., 2013). Las características de no linealidad y no estacionariedad sugieren que tanto los métodos no lineales como los aplicados en ventanas de tiempo deslizables podrían ser los más adecuados para un análisis confiable del registro (Micheloyannis et al. (1998a)).

Los registros electroencefalográficos contienen señales que no provienen de la actividad cerebral basal y que pueden dificultar la interpretación del registro. Estas señales son llamadas artefactos y se clasifican de acuerdo a su origen en fisiológicos (parpadeos, movimientos oculares, de extremidades y cabeza) y externos, que pueden ser debidos al mal colocamiento de los electrodos o a la interferencia de campos eléctricos de aparatos presentes en el lugar en donde se realiza el registro.

Algunos de los métodos más comunmente utilizados para eliminar artefactos en electroencefalogramas son análisis de componentes independientes (ICA, por sus siglas en inglés), wavelets, filtros lineales y análisis de componentes principales (PCA, por sus siglas en inglés). Estudios sugieren que aplicar combinaciones de estos métodos es una forma más efectiva de eliminar artefactos (Patil and Biradar, 2016).

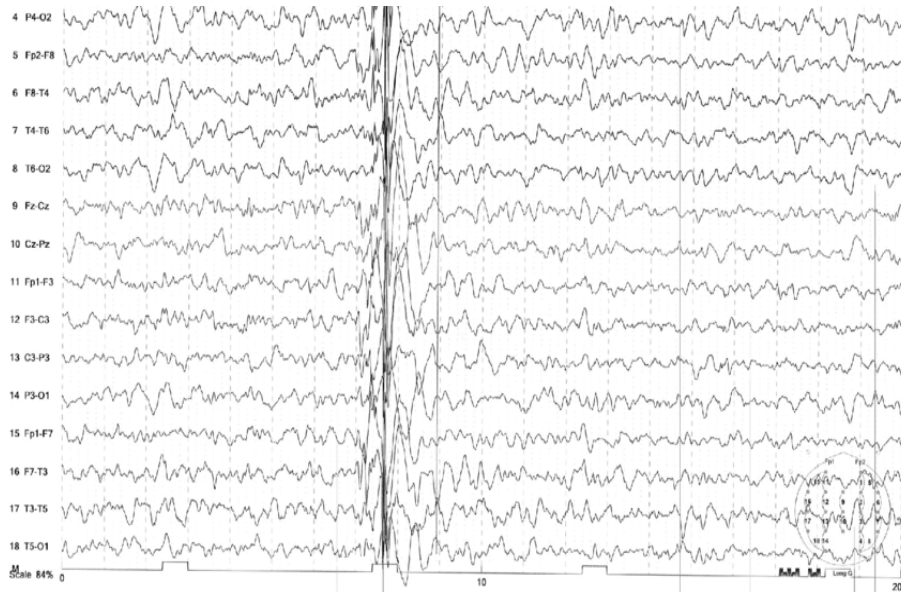


Figura 1.6: *Ejemplo de EEG de un paciente con epilepsia que presenta descargas punta y onda generalizadas, es posible apreciar la actividad anormal alrededor del segundo 8 del registro. Imagen tomada de Pavone et al. (2015).*

Analizar al EEG como una serie de tiempo permite estudiar las manifestaciones de la actividad eléctrica de grandes grupos de neuronas con una alta resolución temporal, esto hace posible detectar cambios de actividad

relacionados con ciertas patologías. En particular, la epilepsia se caracteriza por la presencia abrupta de descargas excesivas de un gran grupo de neuronas, esta actividad anómala puede ser captada en un EEG.

En la figura 1.6 se muestra un ejemplo de un registro de EEG de un paciente epiléptico que presenta un tipo de descarga anormal conocida como *punta y onda*, la cual está relacionada con el CPD (ver Sección 1.2.1). Este patrón de actividad está generalizado, es decir, se presenta en todos los electrodos del registro. Es posible observar que durante el periodo de la actividad punta onda, las diferentes señales se *sincronizan*. Como se discutirá en secciones posteriores, este tipo de anomalía se traduce en un aumento transiente de la correlación estadística entre pares de señales. Esta característica estadística de la manifestación de actividad anormal es la que será explorada en el desarrollo de esta tesis para la identificación de actividad epileptiforme.

## Capítulo 2

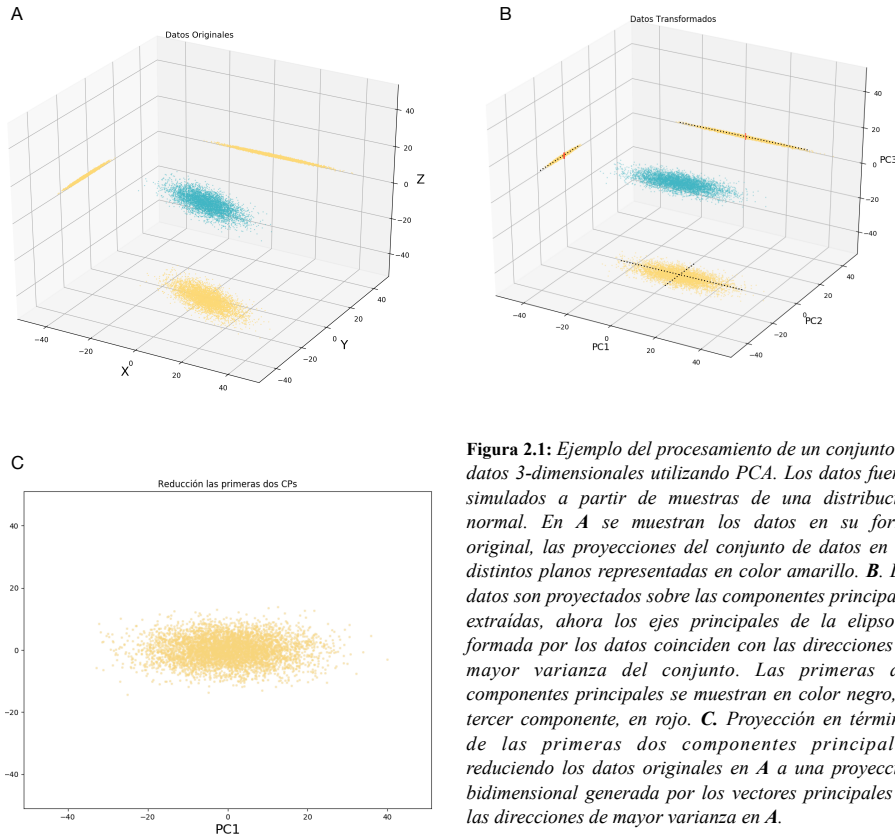
# Análisis de componentes principales

El problema de la alta dimensionalidad en los registros EEG puede ser abordado mediante el uso de métodos de *reducción de dimensionalidad*. Estos métodos funcionan aplicando transformaciones en un conjunto de datos de tal forma que sea posible describir el comportamiento de estos utilizando una menor cantidad de variables.

*Análisis de componentes principales* (PCA) es uno de los métodos de reducción de dimensionalidad más comúnmente usados. Este método se basa en la extracción de vectores mutuamente ortogonales, a lo largo de los cuales se pueden explicar direcciones en las que la varianza de los datos va cambiando, dichos vectores son llamados *componentes principales* (James et al., 2013).

La idea general es muy simple: para una serie de tiempo multivaluada de  $N$  variables, la reducción de dimensionalidad de esta serie consiste en elegir adecuadamente un espacio  $M$ -dimensional sobre el cuál se proyectarán los datos originales. En el caso de PCA, la reducción se enfoca en capturar la mayor cantidad de la varianza del sistema (Fig. 2.1).

En resumen, la aplicación de PCA consiste en realizar un cambio de coordenadas a conjunto  $N$ -dimensional de puntos y restringirlo a un subespacio de dimensión  $M$ , de tal forma que el cambio de coordenadas conserve información acerca de la manera en la que se dispersan los puntos en el espacio de estados original.



**Figura 2.1:** Ejemplo del procesamiento de un conjunto de datos 3-dimensionales utilizando PCA. Los datos fueron simulados a partir de muestras de una distribución normal. En *A* se muestran los datos en su forma original, las proyecciones del conjunto de datos en los distintos planos representadas en color amarillo. *B*. Los datos son proyectados sobre las componentes principales extraídas, ahora los ejes principales de la elipsoide formada por los datos coinciden con las direcciones de mayor varianza del conjunto. Las primeras dos componentes principales se muestran en color negro, la tercer componente, en rojo. *C*. Proyección en términos de las primeras dos componentes principales, reduciendo los datos originales en *A* a una proyección bidimensional generada por los vectores principales en las direcciones de mayor varianza en *A*.

## 2.1. Descripción analítica de PCA

Sea  $\mathbf{X}$  una serie de tiempo multivaluada con  $N$  variables y  $f$  observaciones de cada variable

$$\mathbf{X} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_N \end{bmatrix} \quad (2.1)$$

$\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_N$  denotan los vectores de  $f$  observaciones para cada una de las señales individuales que conforman a la serie de tiempo multivaluada.

Es posible representar la serie  $\mathbf{X}$  en un arreglo matricial, colocando cada una de las señales, en este caso, como renglones y las observaciones de las variables como columnas:

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,f} \\ \vdots & \ddots & \\ x_{N,1} & & x_{N,f} \end{bmatrix}$$

### Matriz de covarianza

La matriz de covarianza  $C$

$$C = \{c_{i,j} \mid c_{i,j} = Cov(\vec{x}_i, \vec{x}_j) : i, j \in \{1, \dots, N\}\}, \quad (2.2)$$

asociada a la serie de tiempo multivaluada  $\mathbf{X}$ , servirá de indicador para saber qué tanto varían las señales contenidas en  $\mathbf{X}$  una con respecto a otra. Cada uno de los elementos de la matriz corresponde a la covarianza entre pares de señales.

### Matrices de transformación

En álgebra lineal, cualquier transformación lineal  $\mathbf{T}$  se puede representar con una matriz  $\hat{\mathbf{T}}$  de tal forma que puede operar en un vector  $\vec{v}$  de  $k$  entradas de la forma

$$\mathbf{T}(\vec{v}) = \hat{\mathbf{T}} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} = \hat{\mathbf{T}} \cdot \vec{v} \quad (2.3)$$

### Eigenvectores y eigenvalores

Note que

$$Cov(X, Y) = Cov(Y, X),$$

de donde  $C$  es una matriz simétrica por lo que es diagonalizable y tiene eigenvectores linealmente independientes. Entonces  $C$  cumple con las condiciones necesarias para ser representada por su descomposición en vectores

propios. Además, los vectores propios correspondientes a  $C$  serán ortogonales.

Recordemos que si  $\vec{u}$  es un vector propio asociado a la matriz de transformación  $\hat{\mathbf{T}}$ , entonces,  $\hat{\mathbf{T}} \cdot \vec{u} = \lambda \vec{u}$ , donde  $\lambda$  es un escalar asociado a  $\vec{u}$  llamado valor propio.

Es posible demostrar que los vectores propios de  $C$  corresponden a las direcciones de crecimiento de la varianza en la serie de tiempo multivaluada original  $\mathbf{X}$ , y que los valores propios asociados a cada uno de ellos representan la varianza propagada a lo largo de tales direcciones. A los primeros  $M$  vectores propios ordenados por el tamaño de sus valores propios en orden descendente se les llama también las  $M$  primeras *componentes principales*.

Una vez ordenados los vectores propios con respecto a sus valores propios en orden descendente, podemos definir una matriz de componentes principales  $P$  colocando los vectores propios como renglones en orden descendente según la magnitud de los valores propios. Esto hará que el primer renglón de arriba a abajo en  $P$  corresponda a la primer componente principal (con el valor propio más alto), mientras que el último renglón de arriba a abajo corresponda a la componente con el valor propio más pequeño.

Denotaremos a las componentes principales asociadas a  $C$  como:  $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_N$  y a los valores propios correspondientes a ellas como:  $\lambda_1, \lambda_2, \dots, \lambda_N$  respectivamente (debido a que  $C$  es una matriz linealmente independiente con  $N$  columnas,  $C$  tiene asociados  $N$  eigenvectores y cada uno tiene  $N$  entradas).

$$P = \begin{bmatrix} \vec{e}_1 \\ \vec{e}_2 \\ \vdots \\ \vec{e}_N \end{bmatrix} = \begin{bmatrix} e_{1,1} & \dots & e_{1,N} \\ \vdots & \ddots & \\ e_{N,1} & & e_{N,N} \end{bmatrix}$$

Nota:  $e_{i,j}$  representa la entrada  $j$  de la  $i$ -ésima componente principal

Multiplicar  $P$  por  $\mathbf{X}$  es equivalente a realizar un cambio de coordenadas



a los puntos de  $\mathbf{X}$  a un nuevo espacio cuyos ejes corresponden con las direcciones de mayor cambio en la varianza (componentes principales) de los datos (Fig. 2.1 B)).

### **Reducción de dimensionalidad**

La reducción de dimensionalidad de PCA consiste en elegir cierto número  $M$  de componentes principales cuyos eigenvalores sean los más altos para generar una matriz de componentes principales *reducida* ( $\tilde{P}$ ), y multiplicarla por la matriz de datos original  $\mathbf{X}$  de tal forma que se obtengan los datos originales reducidos en dimensión con el criterio de las direcciones de mayor varianza. Denotaremos a los datos transformados como

$$\tilde{X} = \tilde{P} \cdot X.$$

Así, el espacio en donde se proyectarán los puntos de  $\mathbf{X}$  será de dimensión  $M$ , es decir, el número de componentes principales que se decida retener.

Es posible obtener la porción de la varianza total (a lo largo de todas las componentes principales) que es explicada por las primeras  $M$  componentes principales ( $V_{E_M}$ )

$$V_{E_M} = \frac{\sum_{i=1}^M \lambda_i}{\sum_{i=1}^N \lambda_i}, \quad (2.4)$$

donde  $N$  es el número de dimensiones del conjunto de datos original,  $M$  es el número de componentes (número de dimensiones) a las que se redujo el conjunto de datos y  $\lambda_i$  es el eigenvalor asociado a la  $i$ -ésima componente principal.

#### **2.1.1. Criterios de reducción de dimensionalidad en PCA**

En general no hay una regla que indique el número de dimensiones a las que es adecuado reducir el conjunto de datos pues esto depende de las características del conjunto de datos que se esté estudiando y del tipo de análisis que se desee realizar. Sin embargo en la práctica existen dos criterios comúnmente utilizados y aceptados. El primero consiste en elegir el número de componentes en función de la varianza explicada ( $V_E$ ) y mantener el número de componentes que expliquen cierto porcentaje de la varianza total. Con la finalidad de mantener la mayor porción de información original en términos de la reducción de dimensiones, en la práctica es deseable explicar

alrededor del 90 % de la varianza total de los datos. Sin embargo, ese porcentaje siempre es arbitrario, pues depende en gran parte de las características de los datos que se estén analizando y comúnmente es posible trabajar con varianzas explicadas mucho menores para obtener resultados robustos.

El segundo método, llamado *scree plot* (James et al., 2013) consiste en realizar una gráfica de las componentes contra su respectivo eigenvalor, de tal forma que el punto en el cual la pendiente de la curva comience a disminuir claramente (correspondiente a la separación entre eigenvalores bajos y altos) funcione como un indicador del número de componentes principales a mantener.

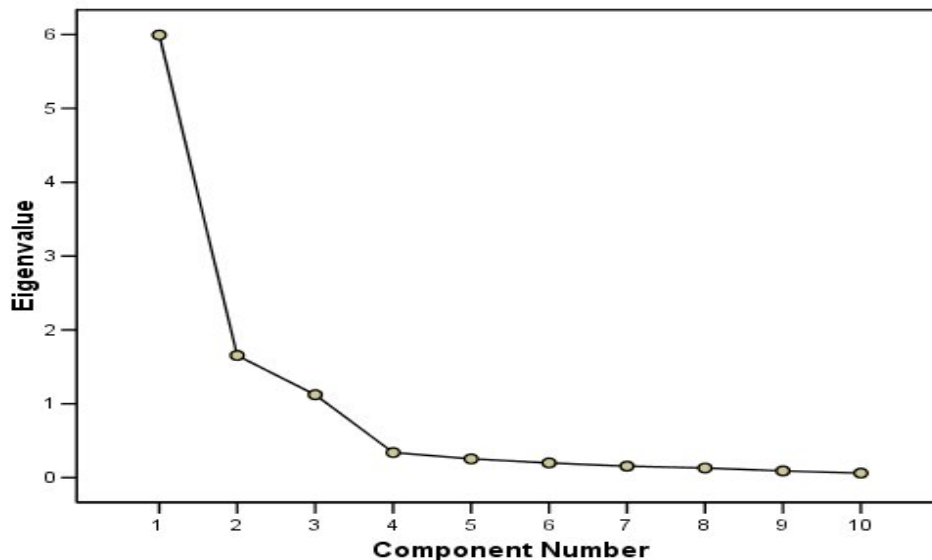


Figura 2.2: Ejemplo de un *scree plot* correspondiente a un conjunto de datos 10-dimensional. El cambio abrupto en la pendiente de esta gráfica corresponde aproximadamente con el segundo o tercer eigenvalor, lo que sugiere que mantener las tres primeras componentes principales podría ser una reducción de dimensionalidad adecuada. Imagen tomada de IBM SPSS Statistics 26 documentation.

## 2.2. PCA dependiente del tiempo

Los usos más comunes de PCA son la reducción de dimensionalidad y ruido en un conjunto de datos, sin embargo existen de casos en los cuales el

proceso subyacente que genera los datos cambia de forma considerable a lo largo del tiempo, provocando a su vez cambios en las propiedades estadísticas de la serie de tiempo observada. Para detectar este tipo de cambios locales a lo largo del tiempo es muy recomendable aplicar algoritmos que tomen en cuenta la no estacionariedad.

Los algoritmos con base en PCA que toman en cuenta la no estacionariedad de los datos son conocidos como *PCA dependiente del tiempo*. Se han propuesto principalmente tres formas distintas de aplicar PCA dependiente del tiempo (De Ketelaere et al., 2015). Estos algoritmos son PCA dinámico (DPCA), PCA recursivo (RPCA) y PCA en ventanas de tiempo movibles (MWPCA). Cada uno de estos algoritmos presenta sus ventajas y desventajas; escoger el más adecuado dependerá de las características específicas del conjunto de datos que se esté analizando.

En las siguientes subsecciones se incluye una descripción general de los algoritmos de PCA dependientes del tiempo más comúnmente utilizados, mismos que fueron descritos y comparados en la revisión de De Ketelaere et al. (2015).

### 2.2.1. PCA Dinámico

Este método fue propuesto por primera vez por Ku et al. (1995) para abordar problemas relacionados con la supervisión de procesos químicos industriales. A diferencia de PCA estático (sin ventanas de tiempo deslizables), este método considera agregar una versión retrasada de la serie de tiempo además de una versión sin retraso como información para calcular la matriz de covarianza. De esta forma el algoritmo PCA dinámico (DPCA, por sus siglas en inglés) toma en cuenta las posibles autocorrelaciones de la serie de tiempo multivariada. Así, se utiliza la información provista por una ventana de tiempo previa para tomar en cuenta la autocorrelación de la serie de tiempo y sensar posibles cambios en la autocorrelación de la serie (De Ketelaere et al., 2015). Sin embargo, se ha demostrado que aplicar DPCA en un conjunto de datos inevitablemente ocasionará la aparición de correlaciones espurias en la transformación realizada por el algoritmo (Kruger et al., 2004), lo que a su vez puede causar falsas alarmas en los resultados, además, la elección adecuada del parámetro de retraso de la serie de tiempo puede no ser sencilla. Además este parámetro puede hacer más complicada la interpretación de los resultados.

### 2.2.2. PCA Recursivo

PCA recursivo (RPCA, por sus siglas en inglés) es un algoritmo que se basa en asignar pesos a las observaciones en la serie de tiempo para disminuir sistemáticamente la contribución de las observaciones pasadas al momento de extraer las componentes principales (De Ketelaere et al., 2015). Para este algoritmo se debe elegir un parámetro de *olvido*. Para procesos altamente no estacionarios, se han implementado algoritmos RPCA que permiten que el factor de olvido cambie con el tiempo.

Elegir el parámetro de olvido puede ser complicado dependiendo de la estructura de correlación de la serie de tiempo multivaluada que se esté analizando. Además, actualizar los valores del parámetro de olvido y calcular las matrices de covarianza pesadas con respecto a este ocasiona que el tiempo de cómputo aumente.

### 2.2.3. PCA con ventana movable

PCA con ventana movable (MWPCA, por sus siglas en inglés) es un algoritmo que se actualiza en cada punto temporal del conjunto de datos recorriendo una *ventana de tiempo* con un tamaño predefinido. Esta versión de PCA dependiente del tiempo limita su análisis a aquellas observaciones que quedan dentro de la ventana de tiempo deslizada. Debido a que en cada ventana de tiempo este algoritmo aplica un PCA estático, una suposición elemental de MWPCA es considerar a las observaciones pertenecientes a cada ventana como parte de un proceso localmente estacionario.

Una ventaja de aplicar MWPCA con respecto a DPCA y RPCA es que este depende sólo del tamaño de la ventana de tiempo. Sin embargo la elección adecuada de la ventana de tiempo deslizable es un problema no trivial y se han propuesto distintas formas de estimar un tamaño apropiado con base en ciertas propiedades estadísticas resultantes del análisis (Russell et al., 2000). Por otro lado, conocer los rangos de tiempo en los cuales el sistema de interés presenta cambios, puede ayudar a elegir un tamaño adecuado para la ventana de tiempo.

## 2.3. Aplicaciones de PCA al estudio de EEGs

Las principales aplicaciones de PCA dependiente del tiempo están enfocadas a la detección de intrusiones en redes de computadoras (Pacha and

Park, 2007) y en la supervisión estadística de procesos. Es decir, para identificar los momentos en los que un sistema se aleja de un comportamiento estadístico específico (De Ketelaere et al., 2015).

Existen pocos estudios que apliquen este tipo de metodologías para el análisis de electroencefalogramas con epilepsia. La mayor parte de estos aplica PCA dentro de un esquema de trabajo más amplio (Acharya et al., 2012); (Wang et al., 2010).

En general, realizar un análisis de series de tiempo aplicado en ventanas de tiempo es equivalente a modelar la serie de tiempo en cuestión como un proceso localmente estacionario. Es decir, como un conjunto de observaciones secuenciales que presentan estacionareidad local, mientras que las características estadísticas de la serie de tiempo completa varían de forma gradual. Es importante tener en cuenta que en ciertos casos es difícil justificar la aproximación de estacionareidad local, especialmente debido a que no hay una generalización natural de un proceso estacionario a uno no estacionario (Dahlhaus, 2012).

Algunos estudios han expuesto la validez de suponer estacionareidad local en una serie de tiempo multivariada para aplicar MWPCA y así extraer información valiosa de los datos. Ombao (2006) utilizó los datos correspondientes a un EEG de 16 canales con una duración de 126 segundos que presentaba una crisis epiléptica para probar un método de análisis que tenía como base la aplicación de MWPCA en el dominio de las frecuencia de las señales del EEG.

En este trabajo de tesis aplicamos MWPCA en registros electroencefalográficos con diagnóstico de epilepsia con actividad de tipo CPD. A partir de la aplicación de MWPCA se calculó la varianza explicada ( $V_E$ ) por las primeras tres componentes principales en cada una de estas ventanas. Debido a que los episodios epileptiformes se caracterizan por cambios abruptos en los perfiles de actividad de las señales EEG en los que se alternan periodos de alta y baja actividad en los registros, suponemos que sería posible capturar información relacionada a tales cambios de actividad de forma indirecta a través de cambios en los valores de la  $V_E$  en ventanas de tiempo sucesivas.

## Capítulo 3

# Resultados y discusión

En este capítulo se describe a detalle la metodología con la que se analizaron los registros EEG, así como las características de los datos explorados (secs. 3.1 - 3.3). Además este capítulo contiene los resultados del análisis, que fueron divididos de la siguiente forma: en la Sección 3.4 se realiza una descripción de la metodología mostrando además los resultados de analizar un EEG de *prueba*. La Sección 3.5 contiene los resultados del análisis general obtenidos después de aplicar la metodología al resto de los registros, en esta misma sección se discute la relación entre las fluctuaciones *abruptas* de  $V_E$  con la presencia de actividad epileptiforme. Finalmente, en la Sección 3.6, se explora la posibilidad de implementar un clasificador automático de eventos de actividad epileptiforme para cada registro.

### 3.1. Obtención y descripción de los datos

Se analizaron 5 EEGs con duración de 30 minutos que presentaban descargas epilépticas interictales (DEI) tipo punta y onda generalizadas, es decir, con presencia de la actividad epileptiforme en todos los canales. Estos registros anonimizados son correspondientes a mujeres jóvenes con edades de entre 18 y 25 años. La frecuencia de muestreo de los registros fue de 512 Hz, los registros fueron obtenidos aplicando el sistema 10-20 internacional con 20 electrodos referenciados al electrodo A-1 (Fig. 3.2). Agradecemos al gabinete clínico de neurofisiología y sueño PLADI por proporcionarnos los datos analizados.

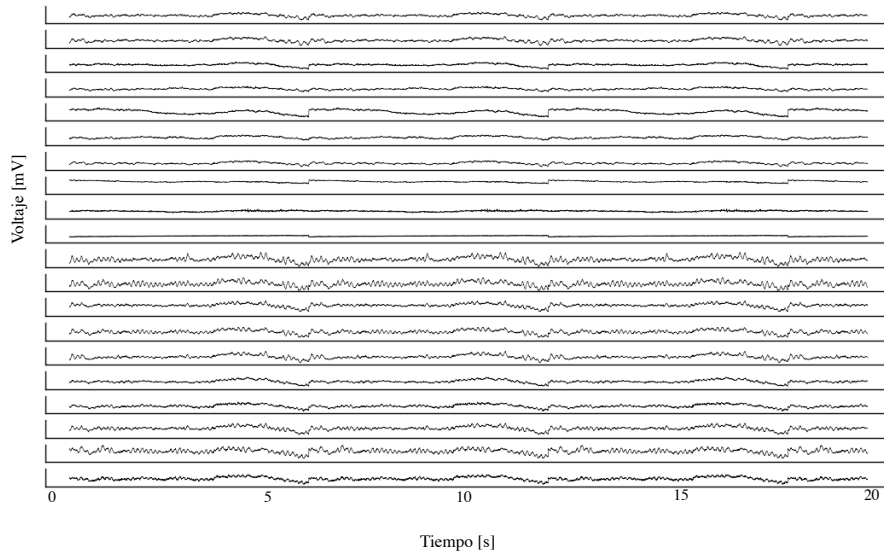


Figura 3.1: Ejemplo de uno de los registros de EEG analizados (20 segundos), los registros fueron realizados utilizando el montaje del 10-20 internacional.

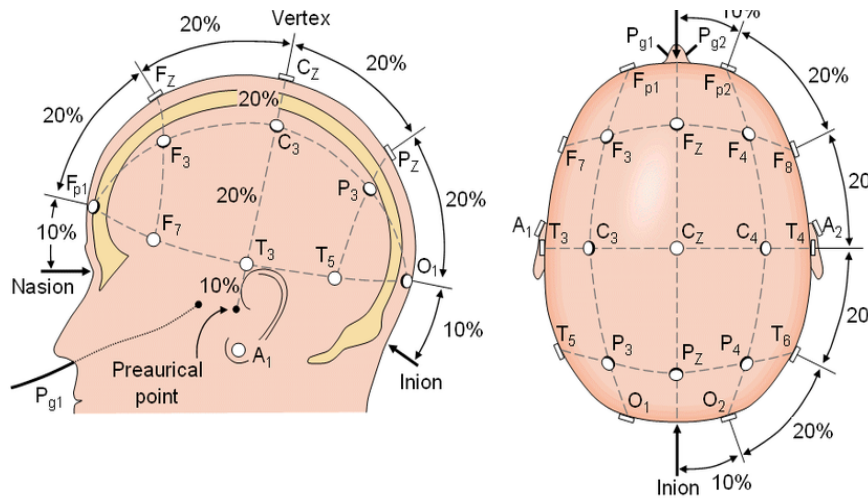


Figura 3.2: Colocación de electrodos del sistema 10-20 internacional. El sistema 10-20 es el internacionalmente estandarizado para la colocación de electrodos en un EEG de superficie. En los registros analizados, el potencial eléctrico de cada señal es referenciado al electrodo A1. Imagen tomada de Olivas et al..

### 3.2. Aplicación de PCA en ventanas de tiempo

Los perfiles de actividad neuronal característicos de la epilepsia corresponden a patrones específicos de la traza del EEG, los cuales son distinguibles de los regímenes de actividad normal. Además, el hecho de que las descargas epilépticas interictales corresponden a la presencia generalizada (en muchos electrodos) de secuencias de alta y baja actividad eléctrica, sugiere que la variación conjunta de las señales en los EEGs también debería cambiar de forma considerable en presencia de actividad epileptiforme, por lo que una herramienta de análisis como PCA, que toma en cuenta la covarianza de un conjunto de observaciones podría ser de gran ayuda para identificar los cambios de actividad relacionados con epilepsia en un EEG.

El análisis presentado aquí tiene como base la aplicación MWPCA (PCA en ventanas de tiempo móviles) con una ventana de tiempo que se mueve en pasos de .5 segundos ( $S = 0.5$  segundos) con tamaños de 1 segundo, 5 segundos y 10 segundos a cada uno de los registros. Es de notar entonces, que los traslapes entre ventanas incrementan en la medida en la que incrementa el tamaño de la ventana.

A partir de las exploraciones iniciales a los EEGs, se observó que 3 componentes principales (CPs) eran suficientes para explicar al menos el 70 % de la varianza total en la mayor parte de las ventanas de tiempo. Debido a que este comportamiento fue observado para todos los registros estudiados, se mantuvieron 3 CPs al aplicar PCA en cada ventana de tiempo para obtener la serie de tiempo de varianza explicada ( $V_E$ , ver Ec. 1.3 en la sección 2.1).

La identificación de periodos de actividad epileptiforme, fue realizada a partir del análisis de las series de tiempo de  $V_E$  para cada registro. Es importante mencionar que el esquema de análisis propuesto tiene el potencial de ser aplicable a otro tipo de series de tiempo multivariadas no estacionarias que presenten cambios abruptos en sus formas de variación conjunta.

### 3.3. Descripción del algoritmo

Supongamos que se tiene un registro con un total de  $F$  observaciones a lo largo del tiempo. La primera parte del algoritmo consiste en dividir el registro en diferentes ventanas de tiempo o segmentos sobre los cuales se aplicará PCA. Para esto se siguieron los pasos descritos a continuación:



1. Definir el tamaño de las ventanas de tiempo  $L$ .
2. Definir la cantidad de deslizamiento de cada ventana ( $S$ ). A partir de  $S$ , la cantidad de traslape entre ventanas de tiempo sucesivas está dada por  $L - S$ . Note que para que cada observación temporal del registro sea incluida en alguna de las ventanas de tiempo se debe de cumplir la condición  $S \leq L$ . Además en el caso  $S = L$ , las ventanas no tendrán traslape.
3. Encontrar los índices correspondientes a los inicios de cada ventana de tiempo. Denotaremos a este conjunto de índices como

$$I = \{i_0, i_1, \dots, i_f\}.$$

Por simplicidad tendremos que  $i_0 = 0$ , correspondiente al inicio del registro. Obsérvese que  $i_f + L \geq F$ . Por tanto, es importante para fines de cómputo escoger  $S$  tal que la igualdad se cumpla, o para que el tamaño de la última ventana no sea demasiado pequeño.

Debido a que el inicio de cada ventana está dado por la cantidad de deslizamiento  $S$  con respecto al punto inicial de la ventana anterior, se tiene simplemente que

$$i_0 = 0, i_1 = S, i_2 = 2S, \dots, i_f = fS.$$

Donde  $f$  es el número de ventanas de tiempo.

4. Encontrar los índices que determinan el final de cada ventana de tiempo. Debido a que el tamaño ( $L$ ) de cada segmento ya fue elegido, basta con sumarle  $L - 1$  a cada índice del conjunto  $I$  para obtener los índices correspondientes al final de cada una de las ventanas de tiempo.

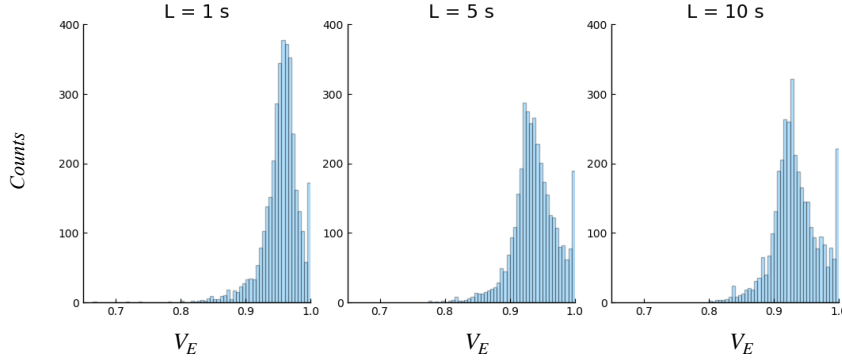


Figura 3.3: *Distribuciones representativas de los valores de  $V_E$  por tres componentes principales en uno de los registros. Los valores de  $V_E$  fueron calculados en ventanas de tiempo de  $L = 1, 5$  y  $10$  segundos respectivamente, manteniendo  $S = 0.5$  segundos. Es posible apreciar que independientemente de si  $L = 1, 5$  ó  $10$  seg, tres componentes principales son suficientes para recuperar más del 80 % de la varianza total en cada ventana de tiempo.*

5. Con las ventanas de tiempo delimitadas, se procede a aplicar PCA a cada una de ellas, reduciendo el registro a tres dimensiones, y posteriormente, se calcula el valor de  $V_E$  en cada ventana de tiempo. Se determinó reducir el registro en cada ventana a tres dimensiones a partir de la observación de que tres componentes principales eran suficientes para recuperar más del 70 % de la varianza total en la mayor parte de las ventanas de tiempo (Fig. 3.3) para todos los registros analizados.

### Validación del algoritmo por un neurólogo experto

Finalmente se comparó el comportamiento de las series de tiempo de  $V_E$  obtenidas para cada paciente con la calificación de un neurólogo a los registros. La calificación del neurólogo consistió en la identificación de los intervalos de 30 seg a lo largo del registro dentro de los cuales se presentaron episodios de actividad epileptiforme con una duración mayor a 400 ms.

Existen perfiles de actividad epileptiforme que pueden tener una duración menor a 400 ms. Sin embargo, debido su corta duración estos biomarcadores son más difíciles de detectar. Por esta razón el análisis se enfocó a la detección de anomalías de duración mayor o igual a 400 ms, un rango similar al de la actividad del PDS.

### 3.4. Descripción del análisis

Después de exploraciones exhaustivas en distintos EEGs, se encontró que las series de tiempo de  $V_E$  presentaban deflexiones abruptas que alcanzaban mínimos locales fácilmente distinguibles, en ventanas de tiempo coincidentes con los intervalos en donde el médico neurofisiólogo había detectado actividad epileptiforme. Debido a esto, la metodología empleada se enfocó en el análisis de las series de tiempo de  $V_E$ ; específicamente en aquellas secciones de la serie de tiempo de  $V_E$  que mostraran cambios abruptos en sus valores. La idea principal es facilitar la identificación automática de regiones en el registro donde pudiera haber actividad epileptiforme, para facilitar la labor del médico experto, y restringir, de ser posible, el tiempo de revisión de los registros, que pueden ser largos.

En esta sección se muestran los resultados de aplicar la metodología de MWPCA censando la varianza explicada ( $V_E$ , ver Sección 1.7.2) por las primeras tres componentes principales en cada ventana tiempo a un sólo registro (eeg\_1).

#### 3.4.1. Análisis de la varianza explicada

La varianza explicada por las primeras 3 componentes a las que se redujeron los registros fue obtenida aplicando la ec. 2.1 (sección 2.1) a cada ventana de tiempo en la que se segmentaron los EEGs. Lo anterior generó una serie de tiempo de  $V_E$  (Fig. 3.4 A). A partir de las series de tiempo  $V_E$  se construyeron visualizaciones de mapeos de color (Fig. 3.4 B) para identificar visualmente los perfiles de fluctuación de  $V_E$ . Para la escala de color de esta visualización se tomaron en cuenta tres cifras significativas de los valores de  $V_E$  obtenidos por el algoritmo.

Para finalizar el análisis se comparó la presencia de fluctuaciones *abruptas* en la serie de tiempo de  $V_E$  con la actividad epileptiforme que fue detectada por un médico neurofisiólogo (Fig. 3.2), donde un cambio fue considerado como *abrupto* después de cuantificar y encontrar de forma sistemática valores de la tasa de cambio de  $V_E$  que fueran significativamente grandes en magnitud. Como se explicará a detalle más adelante, la significancia fue definida en términos de cuantiles para la distribución de valores de la tasa de cambio de  $V_E$ .

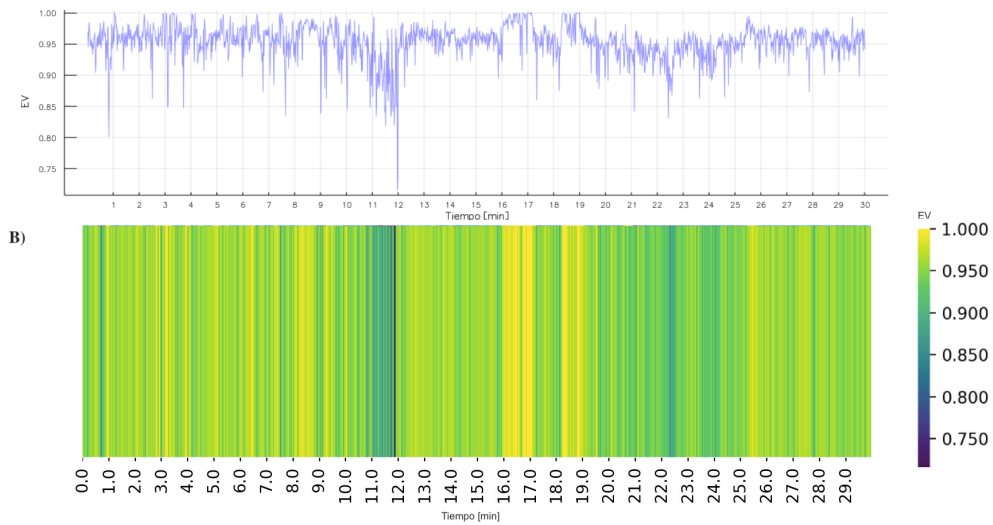


Figura 3.4: En A se muestra un ejemplo de la serie de tiempo de  $V_E$  obtenida para el registro *eeg\_1* mediante MWPCA reduciendo la dimensión a 3 componentes principales. En este ejemplo se utilizaron ventanas de tiempo con un tamaño de 1 segundo y sin traslape ( $S = 1$  seg). En B se muestra un mapeo en escala de colores de la serie de tiempo de  $V_E$ , las fluctuaciones temporales de la serie se aprecian de forma más sencilla como la aparición secuencial de colores altamente contrastantes que representan valores bajos de  $V_E$ .

Las series de tiempo de  $V_E$  para el EEG de prueba (*eeg\_01*) son mostradas en las figuras 3.5 A, B y C. Estas series de tiempo fueron construidas a partir de aplicar MWPCA en ventanas de tiempo de 1, 5 y 10 segundos respectivamente, y con un deslizamiento de la ventana de tiempo en pasos de .5 segundos ( $S = .5$  segundos).

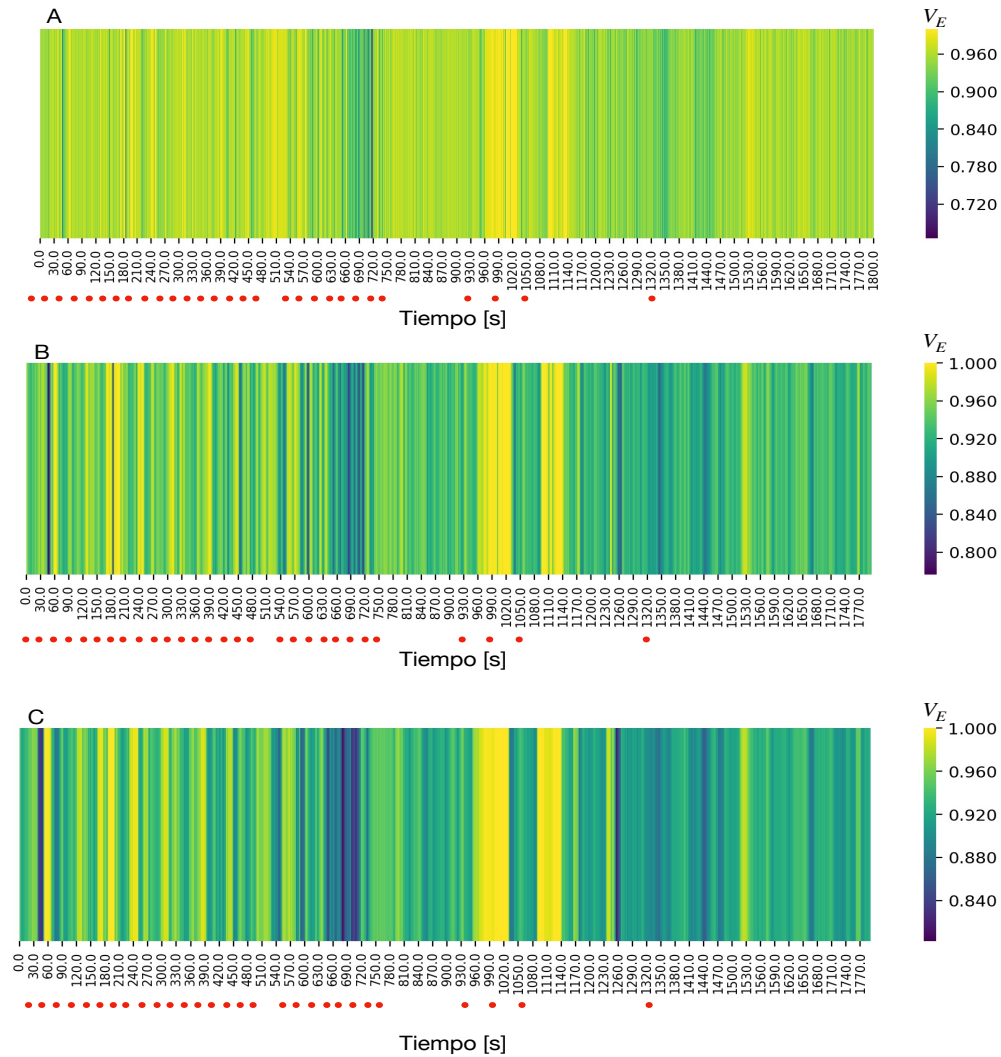


Figura 3.5: Series de tiempo de  $V_E$  mapeadas a una escala de color para el registro eeg\_1. Las figuras A, B y C muestran las series de tiempo  $V_E$  obtenidas en ventanas de tiempo de 1s, 5s y 10s respectivamente. En la escala temporal, los números con un punto en color rojo debajo de ellos corresponden a los intervalos de tiempo de treinta segundos en los cuales el médico neurofisiólogo identificó la presencia de actividad epileptiforme.

El análisis de la serie de tiempo de  $V_E$  aplicado al registro eeg\_1 muestra una relación entre la aparición de fluctuaciones abruptas en la serie de tiempo de  $V_E$  y la presencia de actividad epileptiforme identificada por el neurólogo (imagen superior en la Fig. 3.6). Esta relación puede apreciarse con la coincidencia de la aparición secuencial de colores contrastantes (verde-amarillo-verde) con la actividad anómala detectada (puntos rojos debajo de la escala de tiempo). Para este registro, las fluctuaciones de  $V_E$  son más fáciles de apreciar visualmente en las series de tiempo construidas a partir de ventanas de tiempo de 5 segundos y 10 segundos (Fig. 3.5 B y C), que en la serie construida en ventanas de tiempo de un segundo (Fig. 3.5 A), que muestra una distribución más uniforme en sus valores.

### 3.5. Análisis de fluctuaciones de $V_E$

En esta sección se muestran los resultados de obtener las series de tiempo de  $V_E$  para los registros EEG restantes y comparar con la calificación del neurofisiólogo. Es posible apreciar que existen efectivamente fluctuaciones abruptas en  $V_E$  que corresponden con la presencia de actividad epileptiforme (Figs. 3.5 y 3.5 a 3.8). Estas coincidencias pueden ser apreciadas de forma más clara en las series de tiempo de  $V_E$  con ventanas de  $L = 10$  s (Gráfico C de las Figs. 3.5 y 3.5 a 3.8).

Aunque es posible identificar una relación entre la presencia de actividad epileptiforme tipo PDS (ver Sección 1.2.1) y la presencia de fluctuaciones abruptas en  $V_E$  (Fig 3.8), estas relaciones no son igual de evidentes en todos los registros estudiados (menos claras en eeg\_2 y eeg\_3) en los cuales se presentan fluctuaciones de  $V_E$  que no corresponden con la presencia de actividad epileptiforme en los registros (falsos positivos). Esto puede deberse a que la epilepsia es una enfermedad cuya actividad se presenta en función de una gran diversidad de factores que incluyen la edad del paciente, su alimentación, la localización y características del foco epiléptico, entre otros. Naturalmente, lo anterior da lugar a que los parámetros de MWPCA (como el tamaño de la ventana de tiempo y el número de componentes principales a mantener) que optimizan la detección de candidatos a eventos actividad epileptiforme sean dependientes de las características específicas de cada paciente. Por otro lado, no descartamos que la metodología propuesta también presente sesgos de detección al representar en espacios lineales la serie de tiempo del EEG, que se caracteriza por ser no lineal.

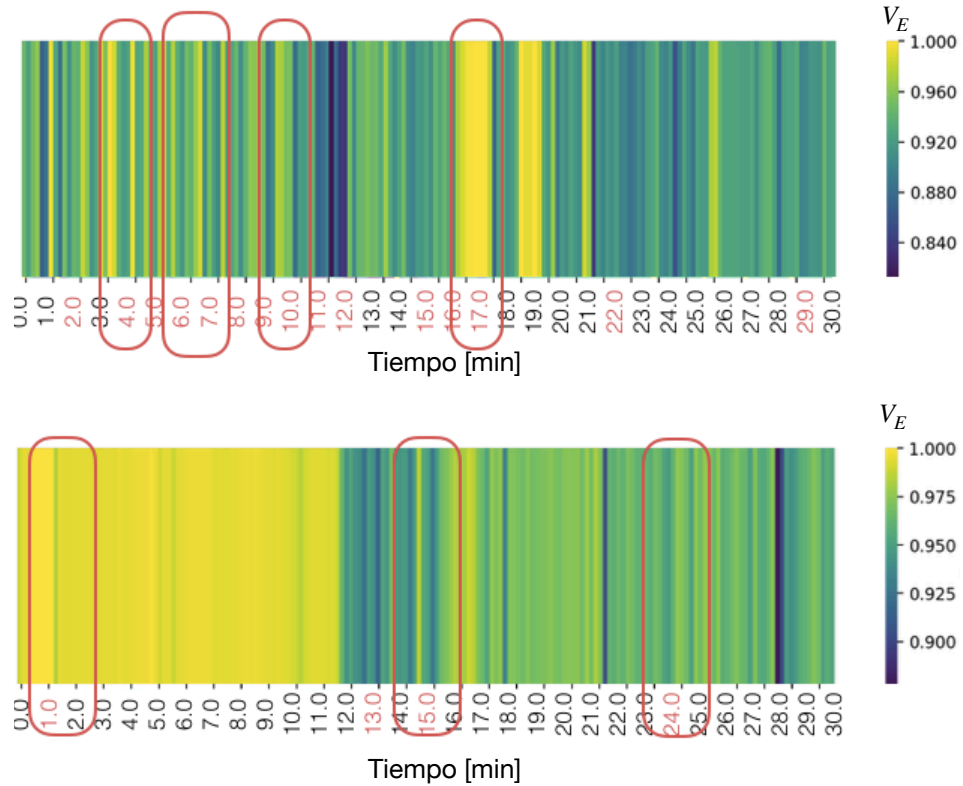


Figura 3.6: Comparación de las series de tiempo de  $V_E$  de los registros *eeg\_1* (arriba) y *eeg\_4* (abajo) con las calificaciones del neurólogo. Los intervalos de 30 segundos en los que el neurólogo identificó actividad epileptiforme se codificaron con números en color rojo. En ambas series de tiempo es posible observar que existen fluctuaciones abruptas en los valores de  $V_E$  (aparición secuencial de colores altamente contrastantes) que son coincidentes con la presencia de descargas epilépticas interictales (DEI) estas coincidencias están señaladas con secciones de las series de tiempo rodeadas por color rojo.

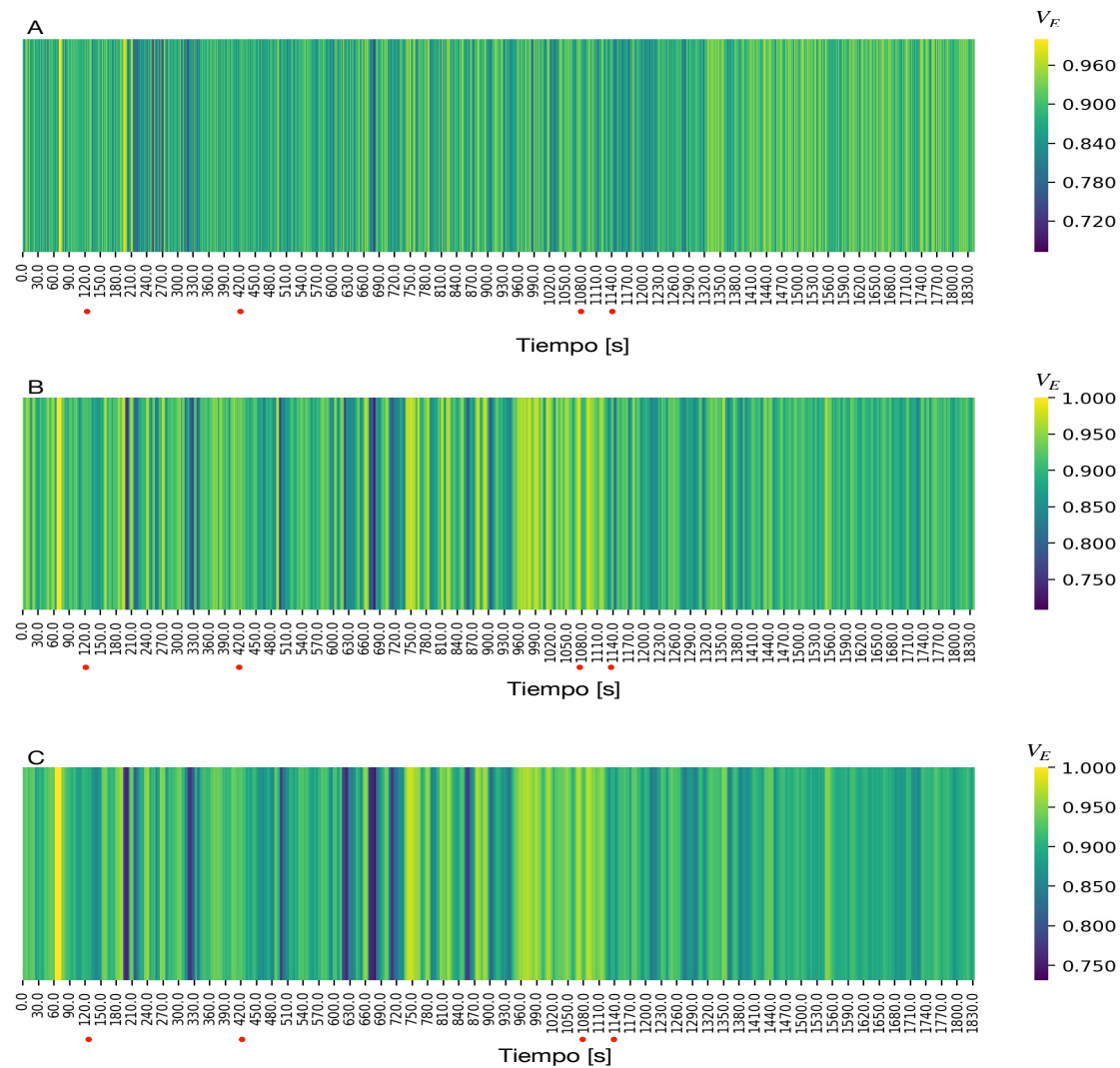


Figura 3.7: Series de tiempo de  $V_E$  para el registro *eeg\_2*. La presencia de actividad epileptiforme en coincidencia con cambios abruptos en la serie de tiempo de  $V_E$  es menos evidente que en el registro *eeg\_1*, siendo posiblemente alrededor del segundo 1080, el único evento detectado. Este decremento en el desempeño de la metodología de detección propuesta puede deberse a que el registro *eeg\_2* se caracteriza por presentar pocos episodios de actividad epileptiforme (6 detectados por el neurólogo), los cuales tienen una duración muy corta (menor a 600 ms).



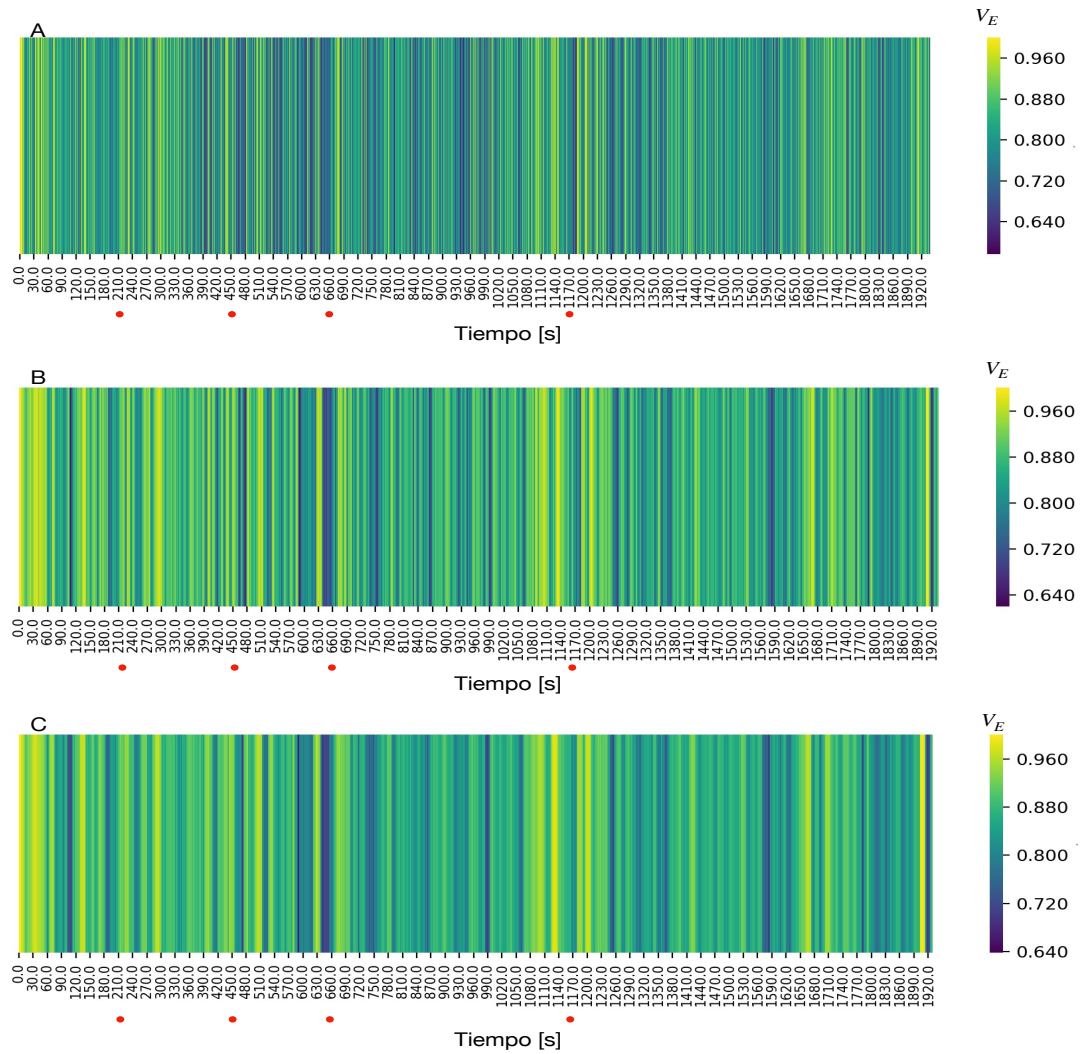


Figura 3.8: Series de tiempo de  $V_E$  para el registro *eeg\_3*. De nuevo, es más sencillo apreciar las fluctuaciones de la varianza explicada en las series de tiempo generadas con ventanas de  $L = 5s$  y  $L = 10s$ . En la serie de tiempo de  $V_E$  con  $L = 10$  segundos se pueden identificar fluctuaciones abruptas relacionadas con patrones epileptiformes en los segundos 660 y 1170, sin embargo, la aparición de estas coincidencias no es tan evidente como en otros casos de detección.

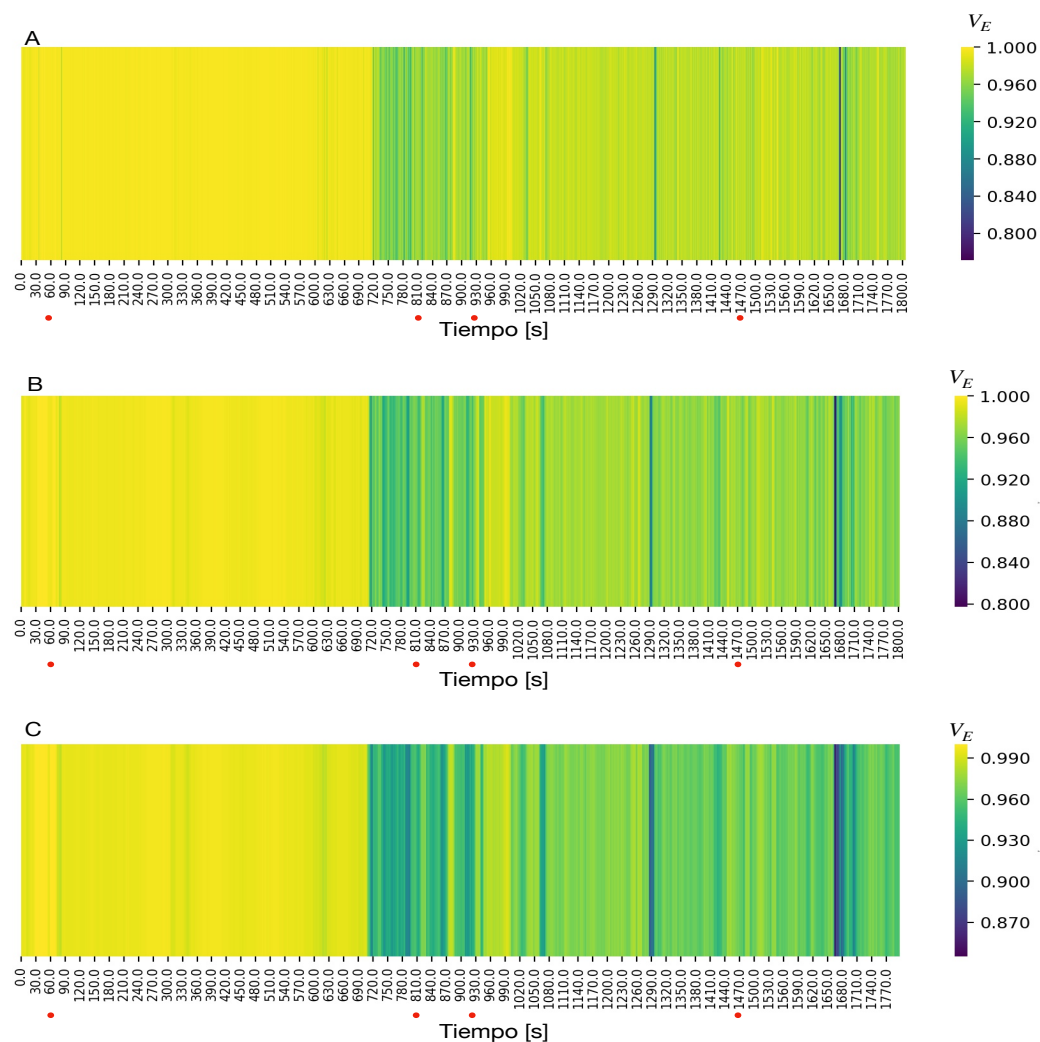


Figura 3.9: Series de tiempo de  $V_E$  para el registro *eeg\_4*. Destaca que en al menos el primer tercio del registro, los valores de  $V_E$  capturados son cercanos a 1.0. Por otro lado, específicamente en C, es posible identificar ciertas fluctuaciones en la serie de tiempo de  $V_E$  que corresponden con la presencia de eventos epileptiformes detectados por el neurólogo (segundos 810 y 930 en C). De nuevo, las fluctuaciones abruptas en  $V_E$  para ventanas de  $L = 1$  seg son más difíciles de visualizar.

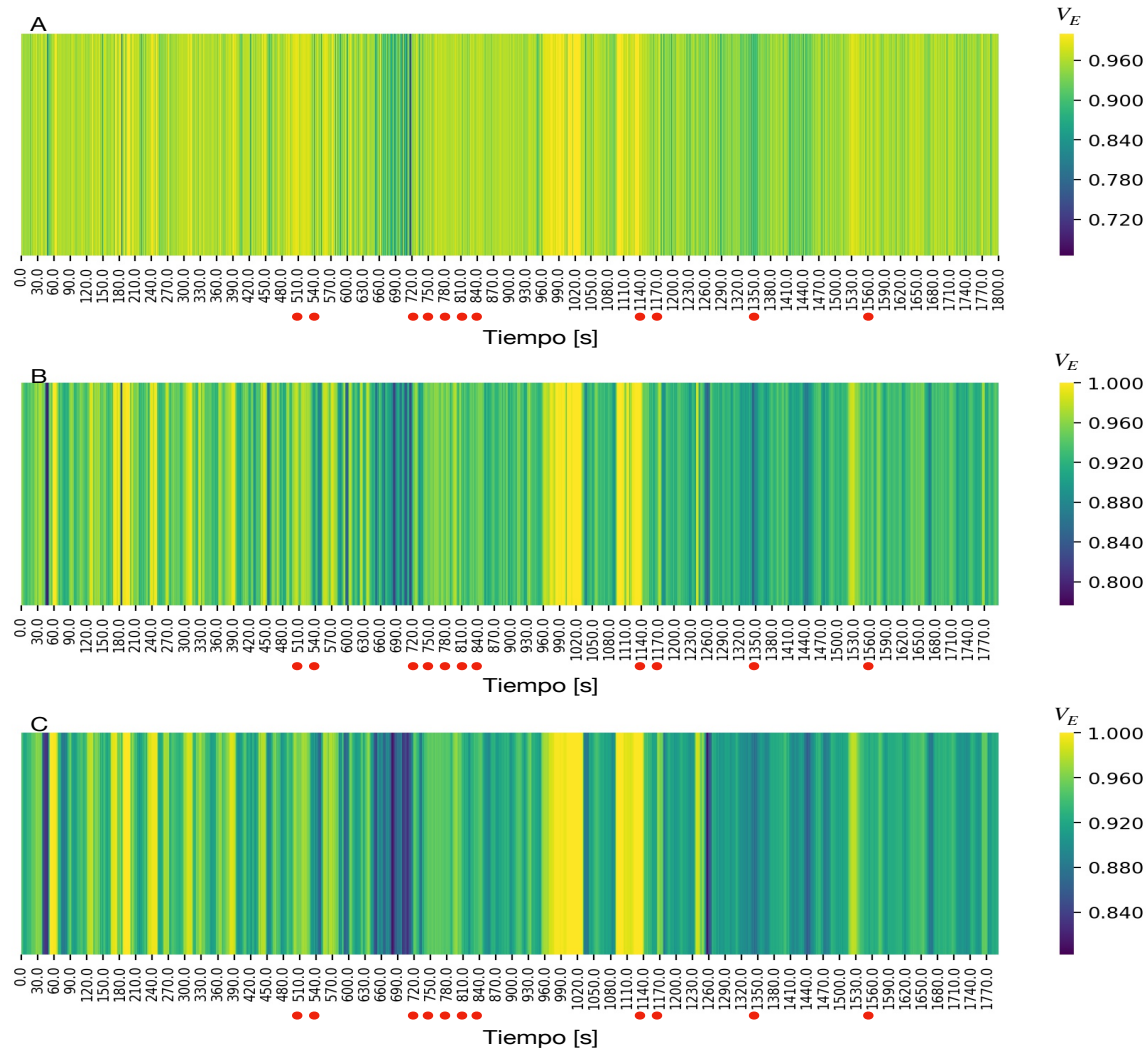


Figura 3.10: Series de tiempo de  $V_E$  para el registro eeg\_5. La serie de tiempo de  $V_E$  reconstruida con ventanas de tiempo de 10 segundos muestra claramente cambios abruptos en los valores de  $V_E$  que coinciden con la actividad epiléptica detectada por el neurólogo (ver C) en los segundos: 510, 720-840, 1170 y 1350). Sin embargo, el hecho de que se presenten muchas fluctuaciones abruptas que no están relacionadas con la presencia de actividad epiléptica (ver segundos 60 y 300-420 de C) indica que el método de análisis puede dar lugar a un número considerable de falsos positivos.

El análisis cualitativo presentado también evidencia la importancia de elegir un tamaño adecuado en las ventanas de tiempo en las que se aplica MWPCA. Proponemos que una elección adecuada del tamaño de la ventana de análisis debe considerar la duración promedio de la anomalía que se espera detectar en los datos. En este caso, el cambio paroxístico en la depolarización (CPD), el cual tiene una duración de aproximadamente 500 ms.

La comparación de la detección de actividad epileptiforme en términos del tamaño de la ventana de tiempo para los registros estudiados sugiere que, al menos ante una inspección ocular de las series de tiempo, es más sencillo identificar fluctuaciones abruptas de  $V_E$  cuando se utilizan ventanas de tiempo de 5s y 10s. Los patrones de fluctuación no son visualmente claros en las series de tiempo de  $V_E$  obtenidas en ventanas de tiempo de 1s.

### 3.6. Clasificador automático de eventos

Finalmente fueron exploradas las ventajas de implementar un clasificador automático de eventos posiblemente relacionados con la presencia de actividad epileptiforme en los registros. En esta sección se muestran resultados preliminares de aplicar el clasificador propuesto a la serie de tiempo de  $V_E$  correspondiente al registro eeg\_1 usando como parámetros  $L = 10$  segundos y  $S = .5$  segundos (fig. 3.5 C). Debido a que observaciones previas de este trabajo nos sugirieron que la presencia de actividad epileptiforme (clasificada así por el médico neurólogo) está relacionada con la aparición de cambios abruptos en los valores de  $V_E$  (Fig. 3.6), las clasificaciones automáticas fueron realizadas en función de los valores de la tasa cambio de  $V_E$ . Para lo anterior se calcularon las diferencias entre valores consecutivos de la serie de tiempo de  $V_E$  y se dividió cada una de las diferencias resultantes entre  $S$ , la cantidad de deslizamiento de las ventanas. Denotaremos a esta nueva serie de tiempo como  $\Delta V_E$  (Fig 3.11).

Para clasificar eventos se dividió la en la serie de tiempo de  $\Delta V_E$  en segmentos de 140 puntos cada uno (equivalentes a 70 segundos, pues  $S = .5$  segundos). Se seleccionaron sólo los valores de  $\Delta V_E$  que estuvieran dos desviaciones estándar sobre o debajo de la media de cada uno de los segmentos (fig. 3.12 A). Esto es equivalente a extraer intervalos temporales durante los cuales se presentan aumentos y decrementos abruptos en los valores de  $V_E$ .

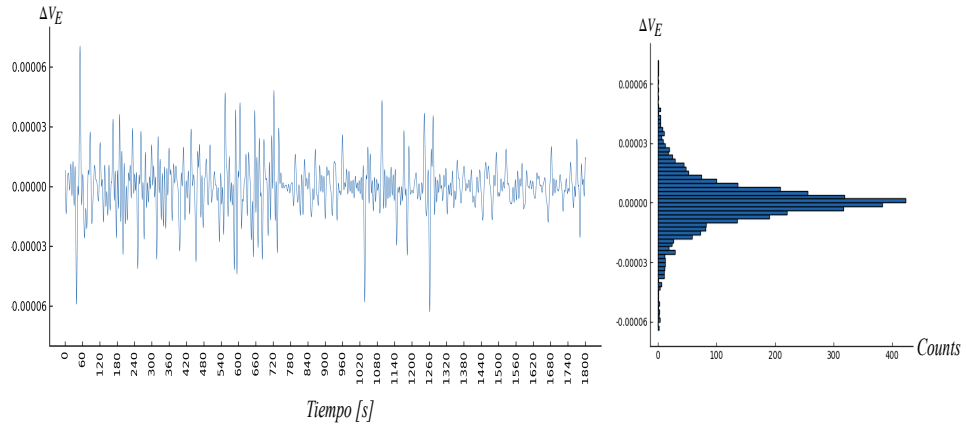


Figura 3.11: Serie de tiempo de  $\Delta V_E$  obtenida para el eeg\_1 a partir de la serie de tiempo de  $V_E$  usando una ventana de un tamaño de 10 segundos que se desplaza en pasos de 0.5 segundos. Los resultados del clasificador automático de eventos de actividad epileptiforme fueron determinados a partir del análisis de esta serie de tiempo y de la distribución de sus valores.

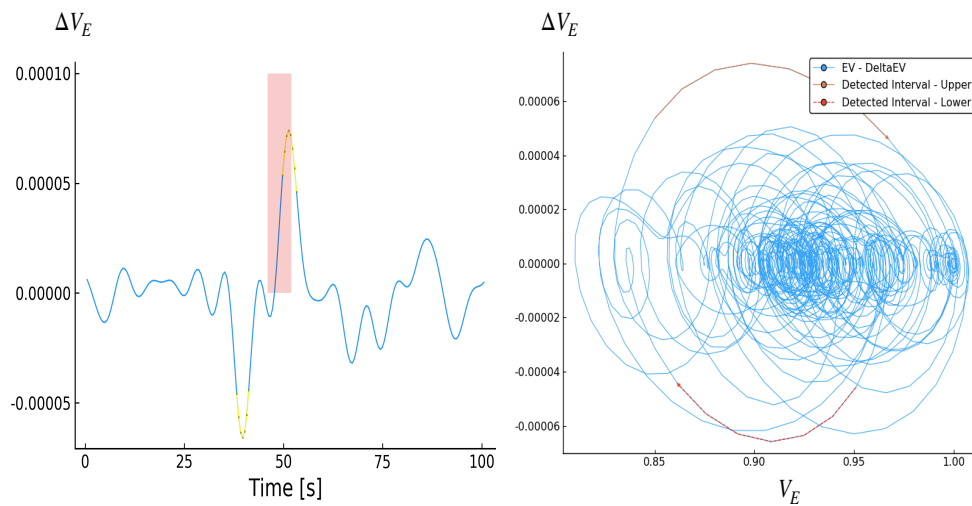
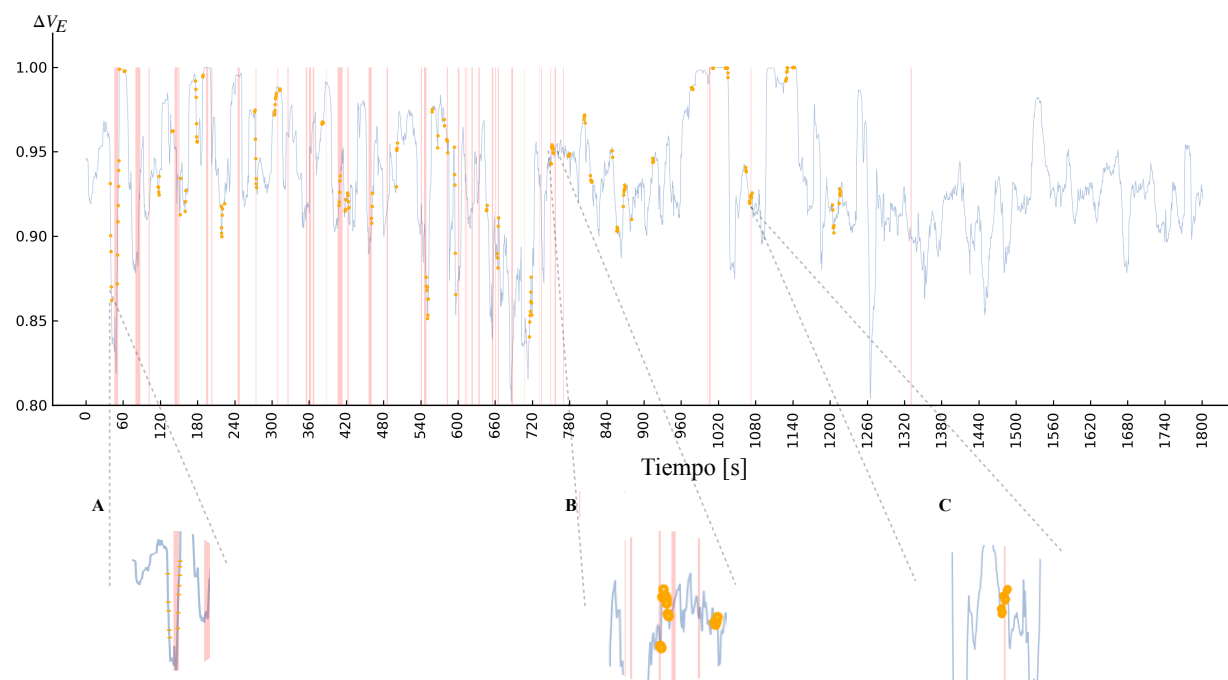


Figura 3.12: A. Ejemplo de puntos clasificados como candidatos a eventos de actividad epileptiforme para el primer segmento de 140 puntos de la serie de tiempo de  $\Delta V_E$  presentada en la fig. 3.11. La coincidencia de la actividad epileptiforme señalada por el neurólogo dentro del segmento (barra roja) con los puntos señalados por el clasificador automático (segmentos amarillos de la curva de  $\Delta V_E$ ) sugiere que la clasificación es acertada dentro de este segmento. B. Retrato fase de las series de tiempo de  $V_E$  y  $\Delta V_E$ . Los segmentos de la curva marcados en color rojo corresponden a los detectados por el clasificador automático, mostrados en la gráfica de la izquierda.

Al aplicar el clasificador a toda la serie de tiempo de  $\Delta V_E$  (fig. 3.13), fue posible observar que los segmentos extraídos por el clasificador automático implementado coinciden con periodos de actividad epileptiforme indicados por el neurólogo en al menos diez ocasiones. Es importante destacar que las figuras 3.13 B y C muestran la existencia de fluctuaciones de la serie de tiempo que no toman valores extremos pero coinciden con la aparición de actividad epileptiforme. Por otro lado, la aparición de puntos amarillos previos pero cercanos a la actividad epileptiforme (barra de color rojo) sugiere que el análisis de esta serie de tiempo podría arrojar indicadores que posiblemente sirvan como alertas tempranas de ciertos perfiles de actividad anormal.



|                    |               | Actividad Real |               |
|--------------------|---------------|----------------|---------------|
|                    |               | Normal         | Epileptiforme |
| Actividad Predicha | Normal        | 3331           | 75            |
|                    | Epileptiforme | 149            | 28            |

Figura 3.13: Visualización del desempeño del clasificador de eventos candidatos a actividad epileptiforme (puntos amarillos). Los episodios de actividad epileptiforme correctamente clasificados son identificables como la coincidencia de regiones con puntos amarillos sobrepuestas a barras rojas, las cuales indican los periodos de actividad epileptiforme señalados por el neurólogo. A. La presencia de regiones con puntos amarillos previos y cercanos a la aparición de actividad epileptiforme sugiere que en ciertos casos  $\Delta V_E$  puede funcionar como alerta temprana a actividad epileptiforme. Los detalles B y C de la figura sugieren que existen fluctuaciones en la serie de tiempo que no toman valores extremos y, al mismo tiempo, corresponden a la presencia de actividad epileptiforme.

## Capítulo 4

# Conclusiones

Si bien el objetivo original de este trabajo de tesis era explorar y comparar diferentes métodos de reducción dimensional en series de tiempo multivaluadas, las observaciones obtenidas después de aplicar MWPCA a EEGs nos ayudó a identificar relaciones entre la varianza explicada por MWPCA y los perfiles de actividad eléctrica registrados en un EEG con actividad epileptiforme. Hemos tratado a la serie de tiempo de  $V_E$  como una serie de tiempo que proporciona información acerca de la forma en que covarían las variables de una serie de tiempo multivaluada. Observar la evolución temporal de  $V_E$  nos permitió identificar la presencia de actividad epileptiforme en algunos de los registros estudiados. Es importante recordar que analizar la serie de tiempo de  $V_E$  es equivalente a extraer patrones de cambio relacionados con la secuencia temporal de las matrices de covarianza de las señales.

Fue posible identificar una relación entre la presencia de actividad epileptiforme y la presencia de fluctuaciones abruptas en la serie de tiempo de  $V_E$  (Fig. 3.6). En gran parte, estas relaciones fueron obtenidas gracias a que (por medio de la aplicación de PCA en ventanas de tiempo) la metodología propuesta en esta tesis toma en cuenta la no estacionariedad característica de los EEGs con epilepsia.

Se observó que la aparición de actividad epileptiforme corresponde con cambios abruptos de los valores de  $V_E$ . Una primera confirmación de esta observación fue realizada a partir de la implementación de un detector estadístico de eventos basado en la localización de cambios repentinos y abruptos en los valores de la serie de tiempo de  $V_E$  (fig. 3.13). En términos de PCA, lo anterior indica que para ventanas de tiempo con actividad epileptiforme



hay una redistribución en los eigenvalores asociados a los eigenvectores de la matriz de covarianza. Esto último también puede ser interpretado como una redistribución en las magnitudes de las direcciones de mayor varianza para los valores de voltaje observados en tales ventanas de tiempo (Fig. 2.1).

Posterior a la disminución transiente en la varianza explicada por tres CPs, se observó la aparición de un aumento súbito en los valores de  $V_E$ , (segmentos encerrados en color rojo en Fig. 3.6), lo que posiblemente correspondería con la componente de poca actividad eléctrica posterior al cambio paroxístico de la depolarización (CDP, ver Sección 1.2.1 y Fig. 1.1). Interpretamos las observaciones anteriores de la siguiente forma: después del CPD la actividad eléctrica disminuye de forma general en todas las señales. Debido a la disminución en la actividad eléctrica registrada, cada señal disminuye drásticamente la dispersión de sus valores alrededor de la media, esto último se refleja en un aumento en la capacidad de capturar una mayor porción de la varianza total de la serie de tiempo usando 3 CPs.

Las relaciones que permitieron identificar la aparición de actividad epileptiforme no son igual de evidentes en todos los registros estudiados. Esto puede ser atribuido a la naturaleza multifactorial de la epilepsia (Sección 1.1) junto con la variación de características entre los pacientes, debido a la limitada cantidad de registros explorados. Por esta misma razón, implementar un sistema de detección automática de actividad anómala en los registros conlleva el riesgo de sobreparametrizar el modelo propuesto.

El análisis de los EEGs sugiere que la precisión de detección es altamente dependiente al menos de:

- La duración e intensidad de los episodios de actividad anormal.
- La cantidad de ruido en los registros.
- Las manifestaciones específicas de la actividad epileptiforme en cada paciente.

La relevancia teórica de este trabajo radica en la propuesta del uso de la varianza explicada por una reducción dimensional como una variable indicadora de cambios abruptos en series de tiempo multivaluadas. La evolución temporal de esta variable, medida en ventanas de tiempo movibles, está relacionada con cambios en las distribuciones estadísticas que presenta la serie de tiempo multivaluada no estacionaria a lo largo de cada ventana de tiempo. En el caso del EEG con actividad epileptiforme, los resultados de esta

tesis nos sugieren que los cambios abruptos en valores de la varianza explicada corresponden a la aparición de actividad epileptiforme. Sin embargo hemos encontrado la existencia de fluctuaciones de la serie de tiempo de  $V_E$  que no toman valores extremos pero parecen corresponder con la aparición de actividad epileptiforme. Consideramos que este es un fenómeno que, de ser encontrado consistentemente en otros EEGs con epilepsia, merecería la pena un estudio a mayor profundidad.

Es importante destacar las ventajas del modelo estadístico presentado, ya que, a diferencia de los usados en otros estudios, ofrece una interpretación de resultados sencilla relacionada con los perfiles de actividad eléctrica que caracterizan a la epilepsia. Además la metodología presentada tiene la capacidad de detectar actividad epileptiforme interictal con duraciones de hasta 400 ms como mínimo, por lo que el modelo podría ser adecuado para detectar crisis epilépticas clínicas.

El programa que se utilizó para estudiar los EEGs fue implementado en Julia, un lenguaje de programación de alto rendimiento, esto permitió extraer las características necesarias realizar el análisis completo de cada registro (con 30 minutos de duración) en cuestión de menos de 30 segundos.

Posibles mejoras a esta primera versión del clasificador contemplan: realizar un preprocesamiento adecuado del EEG con la finalidad de eliminar la mayor cantidad de artefactos del registro; realizar el barrido del clasificador en ventanas de tiempo con traslape, y tomar en cuenta la posible existencia de fluctuaciones no extremas en la serie de tiempo que bien puedan estar relacionadas con la aparición de actividad anormal en el registro.

Aunque a partir de la variación de parámetros del clasificador fuese posible lograr un aumento significativo en el desempeño de detección del algoritmo, esto podría dar lugar a una sobreparametrización del modelo. En este sentido, lo más adecuado sería aplicar el clasificador propuesto a un número mayor de EEGs con diagnósticos de epilepsia similares, de tal forma que se disminuya la varianza del grupo de sujetos estudiados. Este es un estudio que se contempla llevar a cabo en los próximos meses.

Los resultados de esta tesis sugieren que las fluctuaciones abruptas de  $V_E$  podrían ser buenos indicadores de actividad anormal asociada a epilepsia. Sin embargo, debido a la limitada cantidad de registros con los que se contó para realizar el análisis, los resultados aquí presentados están pendientes de

confirmación. Actualmente, en conjunto con el Dr Marco Herrera Valdez, se están haciendo esfuerzos para la aplicar esta metodología en una base de registros EEG con un tamaño adecuado que permita validar o descartar estos hallazgos preliminares.

## 4.1. Posibles extensiones

Existen distintas maneras de mejorar y profundizar el estudio de series de tiempo con base en la metodología de detección propuesta, incluyendo:

- Elegir un método de preprocesamiento adecuado para eliminar artefactos (ruido) de los registros. Esto podría ayudar a refinar la detección de actividad anómala.
- Clasificación de resultados en términos de dimensionalidad: manteniendo fijo un valor para la  $V_E$ , ¿cuántos componentes son necesarios para alcanzar ese valor de  $V_E$ , en cada ventana?
- Aplicar el método de detección a un número mayor de EEGs, seleccionados en función de la presencia de trastornos epilépticos con manifestaciones similares.
- Aplicar la metodología a otras series de tiempo multivariadas no estacionarias.

### 4.1.1. Reconstrucción del espacio fase en series de tiempo

Como exploración adicional de los registros, reconstruimos los espacios fase de los mismos con el método de las componentes principales. A partir de esto se identificó que los puntos en los espacios fase reconstruidos se distribuyen de forma cualitativamente distinta en periodos en donde hay caídas abruptas en la serie de tiempo de  $V_E$  (Fig. 4.1), mismos en los que posteriormente el neurólogo identificó actividad epileptiforme. Hemos descrito estas observaciones a mayor detalle en la siguiente sección.

Un acercamiento al estudio de series de tiempo es utilizar herramientas derivadas de la teoría de sistemas dinámicos no lineales (Micheloyannis et al., 1998b). Este tipo de métodos están basados en intentar reconstruir el espacio fase (o espacio de estados) asociado a un sistema dinámico que subyace a la serie de tiempo de interés con el objetivo de identificar y caracterizar

posibles atractores. Entre las características más comúnmente estudiadas de un espacio fase reconstruido se encuentran la dimensión del atractor, la entropía de Kolmogorov y el exponente de Lyapunov. Aunque existen análisis más elaborados que se enfocan en la detección específica de cambios abruptos en una serie de tiempo (Marwan and Kurths, 2015); (Gao et al., 2017).

Aunque existen estudios que argumentan que la reconstrucción de un espacio fase asociado a un EEG puede ser poco confiable (Pritchard and Duke, 1995) y existe una discusión al respecto (Pezard et al., 1999), este tipo de análisis ha arrojado resultados valiosos para el estudio de la actividad cerebral en pacientes con epilepsia, esquizofrenia, enfermedad de Alzheimer y depresión. Como resultado adicional de esta tesis, se utilizó el método propuesto por Abarbanel (2012) para la reconstrucción de espacios fase de series de tiempo multivariadas en términos de las primeras 3 componentes principales.

Los espacios fase reconstruidos mostrados en la figura 4.1 son correspondientes a una secuencia de ventanas de tiempo ( $L = 10$  seg, sin traslape) del registro eeg\_1 que presentan fluctuaciones abruptas de  $V_E$ . Además, dentro de estos mismos segmentos del registro (en las ventanas de tiempo 100 a 102), el neurofisiólogo identificó actividad epileptiforme.

En los espacios fase reconstruidos es posible observar que antes (Fig. 4.1 I) y después de la aparición de actividad epileptiforme (Fig. 4.1 V y VI), los puntos en el espacio fase se muestran más agregados, en un comportamiento que sugiere la presencia de un *atractor*. Por otro lado, en las ventanas de tiempo en donde aparece actividad epileptiforme (Fig. 4.1 II, III y IV), los espacios fase reconstruidos exhiben una estructura apreciablemente distinta y con una mayor dispersión en la distribución de sus puntos (Fig. 4.1).

Los cambios cualitativos en las topologías de los espacios fase de componentes principales ante la presencia de actividad epiléptica sugieren que este tipo de exploración podría ser utilizada para ayudar a detectar actividad anormal en los registros electroencefalográficos.

Una extensión de los resultados aquí presentados podría enfocarse en estudiar a mayor detalle las relaciones no triviales entre las características de los espacios fase reconstruidos y los pesos asociados a las componentes principales en cada ventana de tiempo.

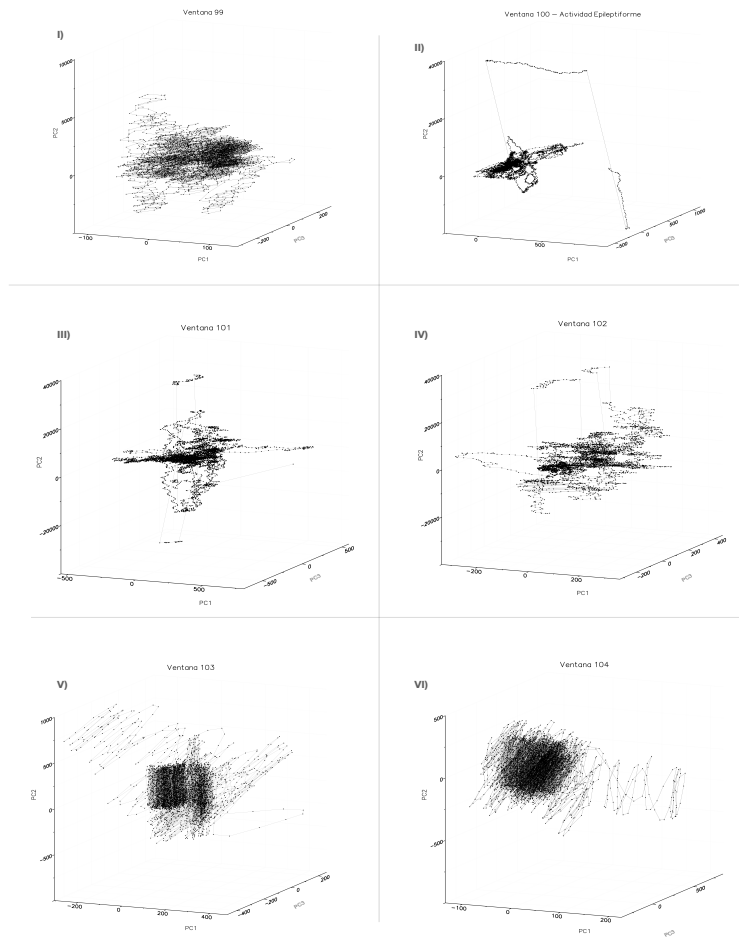


Figura 4.1: *Reconstrucción del espacio fase de Componentes Principales para el registro eeg\_1. Las figuras I - VI muestran la reconstrucción de los espacios fase en distintas ventanas de tiempo con  $L = 10$  segundos y sin traslape en las que se aplicó PCA. Es posible observar que los espacios fase reconstruidos exhiben topologías cualitativamente distintas en las ventanas de tiempo. Específicamente los puntos del espacio fase reconstruidos se muestran más agregados antes (I) y después (V y VI) de la ventanas 100,101 y 102, en las cuales el neurólogo detectó actividad epileptiforme (II, III, IV). Note que en las ventanas que contienen actividad epileptiforme (100,101 y 102), los espacios fase reconstruidos exhiben una mayor dispersión en la distribución de sus puntos.*

## 4.2. Agradecimientos

Muchas gracias a mi familia. A mi padre, a Irene Liu y al Dr. Marco Arieli Herrera Valdez por todo el apoyo y motivación brindada para la realización de este emocionante trabajo. Agradezco también el financiamiento proporcionado por los programas PAPIIT-UNAM IA208618 y PAPIME PE213219 durante la realización de esta tesis y al grupo de trabajo del Dr. Marco Arieli Herrera Valdez por las discusiones que ayudaron al desarrollo de este trabajo.

# Bibliografía

- Henry Abarbanel. *Analysis of observed chaotic data*. Springer Science & Business Media, 2012.
- U Rajendra Acharya, S Vinitha Sree, Ang Peng Chuan Alvin, and Jasjit S Suri. Use of principal component analysis for automatic classification of epileptic eeg activities in wavelet framework. *Expert Systems with Applications*, 39(10):9072–9078, 2012.
- Montserrat et al. Asensio. *Guía práctica de Epilepsia de la comunidad Valenciana*. Springer Healthcare, 2010.
- David A Belsley, Edwin Kuh, and Roy E Welsch. Wiley series in probability and statistics. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, 1980.
- Agustina Garcés Correa, Lorena Orosco, Pablo Diez, and Eric Laciari. Automatic detection of epileptic seizures in long-term eeg records. *Computers in biology and medicine*, 57:66–73, 2015.
- Rainer Dahlhaus. Locally stationary processes. In *Handbook of statistics*, volume 30, pages 351–413. Elsevier, 2012.
- Bart De Ketelaere, Mia Hubert, and Eric Schmitt. Overview of pca-based statistical process-monitoring methods for time-dependent, high-dimensional data. *Journal of Quality Technology*, 47(4):318–335, 2015.
- John S Ebersole. Eeg dipole modeling in complex partial epilepsy. *Brain topography*, 4(2):113–123, 1991.
- Robert S Fisher, Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J Helen Cross, Christian E Elger, Jerome Engel, Lars Forsgren, Jacqueline A French, Mike Glynn, et al. Ilae official report: a practical clinical definition of epilepsy. *Epilepsia*, 55(4):475–482, 2014.

- Zhong-Ke Gao, Michael Small, and Juergen Kurths. Complex network analysis of time series. *EPL (Europhysics Letters)*, 116(5):50001, 2017.
- GBD. Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the global burden of disease study 2016. <https://www.ncbi.nlm.nih.gov/pubmed/30773428>, 2016.
- Gary D Hammer and Stephen J McPhee. *Pathophysiology of Disease: An Introduction to Clinical Medicine 7/E*. McGraw Hill Professional, 2014.
- R Harper, R Sclabassi, and T Estrin. Time series analysis and sleep research. *IEEE Transactions on Automatic Control*, 19(6):932–943, 1974.
- HMG. Epilepsia: Epidemiología. <http://epilepsiaenmexico.com/epidemiologia/>, 2017.
- Kimberly Holland. Epilepsy by the numbers: Facts, statistics, and you. <https://www.healthline.com/health/epilepsy/facts-statistics-infographic>, 2015.
- Leonidas D Iasemidis and J Chris Sackellares. review: Chaos theory and epilepsy. *The Neuroscientist*, 2(2):118–126, 1996.
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Statistical learning. In *An Introduction to Statistical Learning*, pages 15–57. Springer, 2013.
- Herbert Henri Jasper. *Jasper’s basic mechanisms of the epilepsies*, volume 80. OUP USA, 2012.
- JGR Jefferys. Basic mechanisms of epilepsy. <https://www.epilepsysociety.org.uk/sites/default/files/attachments/Chapter03Jeffreys2015.pdf>, 2017.
- Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, A James Hudspeth, et al. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.
- Wlodzimierz Klonowski. Everything you wanted to ask about eeg but were afraid to get the right answer. *Nonlinear Biomedical Physics*, 3(1):2, 2009.
- Uwe Kruger, Yiqi Zhou, and George W Irwin. Improved principal component monitoring of large-scale processes. *Journal of Process Control*, 14(8):879–888, 2004.



- Wenfu Ku, Robert H Storer, and Christos Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1):179–196, 1995.
- Christoph Lossin, Dao W Wang, Thomas H Rhodes, Carlos G Vanoye, and Alfred L George Jr. Molecular basis of an inherited epilepsy. *Neuron*, 34(6):877–884, 2002.
- William W Lytton. Computer modelling of epilepsy. *Nature Reviews Neuroscience*, 9(8):626, 2008.
- Norbert Marwan and Jürgen Kurths. Complex network based techniques to identify extreme events and (sudden) transitions in spatio-temporal systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(9):097609, 2015.
- S Micheloyannis, N Flitzanis, E Papanikolaou, M Bourkas, D Terzakis, S Arvanitis, and CJ Stam. Usefulness of non-linear eeg analysis. *Acta neurologica scandinavica*, 97(1):13–19, 1998a.
- S Micheloyannis, N Flitzanis, E Papanikolaou, M Bourkas, D Terzakis, S Arvanitis, and CJ Stam. Usefulness of non-linear eeg analysis. *Acta neurologica scandinavica*, 97(1):13–19, 1998b.
- NICE. Epilepsies: diagnosis and management. <https://www.nice.org.uk/guidance/cg137>, 2017.
- Carlos Augusto Novo Olivas, Leticia Chacón Gutiérrez, and José Alberto Barradas Bribiesca. Mapeo electroencefalográfico y neurofeedback.
- Hernando Ombao. Time-dependent frequency domain principal components analysis of multichannel non-stationary signals. *Computational statistics & data analysis*, 50(9):2339–2360, 2006.
- Animesh Patcha and Jung-Min Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer networks*, 51(12):3448–3470, 2007.
- Veeresh Patil and Arun Biradar. A survey on methods on artifact removal from eeg. 2016.
- Piero Pavone, Andrea Domenico Praticò, Martino Ruggieri, and Raffaele Falsaperla. Hypomelanosis of ito: a round on the frequency and type of epileptic complications. *Neurological Sciences*, 36(7):1173–1180, 2015.

- Luiz Pessoa. Understanding brain networks and brain organization. *Physics of life reviews*, 11(3):400–435, 2014.
- Laurent Pezard, Jean-Philippe Lachaux, Nitza Thomasson, and Jacques Martinerie. Why bother to spatially embed eeg? comments on pritchard et al., *psychophysiology*, 33, 362–368, 1996. *Psychophysiology*, 36(4):527–531, 1999.
- Walter S Pritchard and Dennis W Duke. Measuring chaos in the brain—a tutorial review of eeg dimension estimation. *Brain and cognition*, 27(3):353–397, 1995.
- Evan L Russell, Leo H Chiang, and Richard D Braatz. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 51(1):81–93, 2000.
- George J Siegel, Bernard W Agranoff, R Wayne Albers, Perry B Molinoff, and Marjorie B Lees. Basic neurochemistry. *Journal of Neuroscience Research*, 39(4):512, 1994.
- SJM Smith. Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2), 2005.
- CJ Stam, JPM Pijn, P Suffczynski, and FH Lopes Da Silva. Dynamics of the human alpha rhythm: evidence for non-linearity? *Clinical Neurophysiology*, 110(10):1801–1813, 1999.
- Cornelis J Stam. Nonlinear dynamical analysis of eeg and meg: review of an emerging field. *Clinical neurophysiology*, 116(10):2266–2301, 2005.
- Joseph I. Sirven Steven C. Schachter, Patricia O. Shafer. Diagnosing epilepsy. <https://www.epilepsy.com/learn/diagnosing-epilepsy>, 2014. Revisado el 28 de Diciembre de 2017.
- Chunmei Wang, Junzhong Zou, Jian Zhang, Min Wang, and Rubin Wang. Feature extraction and recognition of epileptiform activity in eeg by combining pca with apen. *Cognitive neurodynamics*, 4(3):233–240, 2010.
- WebMD. Common epilepsy causes and seizure triggers. <https://www.webmd.com/epilepsy/guide/epilepsy-causes>, 2019.
- WHO. Epilepsy fact sheet. <http://www.WHO.int/mediacentre>, 2017.