



Universidad de Valladolid

FACULTAD DE CIENCIAS

Trabajo Fin de Máster

Máster en Matemáticas

**MÉTODOS PARAMÉTRICOS Y NO PARAMÉTRICOS EN EL ANÁLISIS DE
SEÑALES OSCILATORIAS DE ACTIVIDAD DIARIA, SUEÑO Y RITMO CARDÍACO**

Autor: Miguel Cubero Gómez
Tutoras: Yolanda Larriba, Itziar Fernández

*Dedicado a
mi amigo Javier.*

Agradecimientos

A mis tutoras, Itziar y Yolanda, por aceptar guiarme en este trabajo. Su paciencia y apoyo han sido fundamentales a lo largo de todo el proceso.

A mis padres por su apoyo incondicional mientras combinaba trabajo y estudios. Su comprensión y aliento han sido una fuente constante de motivación.

A mi pareja y a mis amigos por su amor y cariño durante este proceso. Su apoyo emocional ha sido crucial para mantenerme enfocado y positivo.

Resumen

La alteración del ritmo circadiano humano se ha vinculado con diversas enfermedades crónicas, como diabetes, hipertensión y depresión, impulsando una creciente investigación en cronobiología. La tecnología actual permite la recopilación continua de datos biométricos mediante dispositivos portátiles, facilitando el estudio de patrones circadianos y su impacto en la salud. Conjuntos de datos como MMASH ofrecen información detallada sobre la actividad diaria y el sueño. Estos conjuntos recogen variables que presentan patrones oscilatorios a lo largo de las 24 horas del día, y su análisis es esencial para avanzar en la comprensión de la biología circadiana. El trabajo se enfoca en mejorar la comprensión de los ritmos circadianos y desarrollar herramientas precisas para su análisis, abordando las limitaciones de los modelos tradicionales y adaptando nuevas metodologías a las características específicas de los datos circadianos. En concreto, los objetivos de este trabajo son: 1) demostrar la eficacia del modelo FMM en el análisis de señales circadianas obtenidas de la actividad física y la frecuencia cardíaca, y clasificar usuarios según sus parámetros; 2) desarrollar una nueva medida y explorar métodos no paramétricos para analizar la variabilidad de las señales biológicas; y 3) calcular índices no paramétricos y establecer relaciones con los datos del conjunto MMASH.

Abstract

Human circadian rhythm disruption has been linked to various chronic diseases such as diabetes, hypertension, and depression, driving increasing research in chronobiology. Current technology enables the continuous collection of biometric data through wearable devices, facilitating the study of circadian patterns and their impact on health. Datasets like MMASH provide detailed information on daily activity and sleep. These datasets capture variables that exhibit oscillatory patterns throughout the 24-hour day, and their analysis is crucial for advancing the understanding of circadian biology. This work focuses on enhancing the understanding of circadian rhythms and developing accurate tools for their analysis, addressing the limitations of traditional models and adapting new methodologies to the specific characteristics of circadian data. Specifically, the objectives of this work are: 1) to demonstrate the effectiveness of the FMM model in analyzing circadian signals obtained from physical activity and heart rate, and to classify users based on these parameters; 2) to develop a new measure and explore non-parametric methods for analyzing biological signal variability; and 3) to calculate non-parametric indices and establish relationships with the data from the MMASH dataset.

Índice

1	Introducción	15
2	Métodos	18
2.1	Modelos paramétricos	18
2.1.1	Modelo cosinor	19
2.1.2	Modelo FMM	19
2.2	Métodos no paramétricos	20
2.2.1	Índices no paramétricos	21
2.2.2	Análisis de Componentes Principales Funcionales	25
2.2.3	Desarrollo de una nueva medida	28
3	Descripción del conjunto MMASH	30
3.1	Descripción del conjunto de datos	30
3.1.1	Datos de los usuarios (user_info.csv)	31
3.1.2	Datos con respecto al sueño (sleep.csv)	32
3.1.3	Datos de latido (RR.csv)	33
3.1.4	Datos del cuestionario (questionnaire.csv)	33
3.1.5	Datos de actividades (activity.csv)	36
3.1.6	Datos del acelerómetro (actigraph.csv)	37
3.1.7	Datos de las pruebas de saliva (saliva.csv)	38
4	Resultados	39
4.1	Preprocesado	40
4.2	Resultados de los modelos paramétricos	41
4.2.1	Clustering jerárquico con 2 clusters	43
4.2.2	Método K-medias con 5 clusters	45
4.3	Resultados obtenidos de los métodos no paramétricos	51
4.3.1	Resultados del Análisis de Componentes Principales Funcionales (ACPF)	51
4.3.2	Resultados de los índices no paramétricos y cuestionarios	58
5	Conclusiones	68
A	Teoremas del análisis de datos funcionales	69
B	Tablas complementarias	72

Índice de figuras

1	Evolución de las variables de wMvFC para el usuario 3 del conjunto MMASH.	30
2	Frecuencia de los valores de PSQI.	35
3	Frecuencia de las actividades.	37
4	Gráfico de cortisol.	39
5	Gráfico de melatonina.	39
6	Proyección de los clusters obtenidos con el método jerárquico en dos componentes principales. Cada grupo es representado con un color diferente.	44
7	Proyección de los clusters obtenidos con el método K-medias (k=5) en dos componentes principales. Cada grupo es representado con un color diferente.	45
8	(a) Onda FMM obtenida con la mediana de los parámetros del cluster 1. (b) Modelo FMM ajustado para el usuario 2 de dicho cluster. . . .	48
9	(a) Onda FMM obtenida con la mediana de los parámetros del cluster 2. (b) Modelo FMM ajustado para el usuario 13 de dicho cluster. . .	48
10	(a) Onda FMM obtenida con la mediana de los parámetros del cluster 3. (b) Modelo FMM ajustado para el usuario 1 de dicho cluster. . . .	48
11	(a) Onda FMM obtenida con la mediana de los parámetros del cluster 4. (b) Modelo FMM ajustado para el usuario 18 de dicho cluster. . .	49
12	(a) Onda FMM obtenida con la mediana de los parámetros del cluster 5. (b) Modelo FMM ajustado para el usuario 16 de dicho cluster. . .	49
13	Comparativa de los modelos cosinor y FMM para el usuario 1.	50
14	Comparativa de los modelos cosinor y FMM para el usuario 8.	50
15	Función media $\rho(t)$	51
16	Cuatro primeras autofunciones.	52
17	Aproximación ACPF para el usuario 4.	53
18	Ejemplo de DTW para dos series temporales.	54
19	Aproximación para todos los usuarios clasificados en dos grupos con K-medias, DTW y ACPF.	56
20	Aproximación ACPF y ajuste FMM para el usuario 8 del cluster 1. . .	57
21	Aproximación ACPF y ajuste FMM para el usuario 1 del cluster 2. . .	57
22	Melatonina ajustada y Amplitud Absoluta.	61
23	Frecuencia cardiaca por hora para el user_19 con IV=0.08.	63
24	Frecuencia cardiaca por hora para el user_2 con IV=0.39.	63
25	Frecuencia cardiaca por hora para el user_10 con CI=0.03.	65

26	Frecuencia cardiaca por hora para el user_13 con CI=0.81.	65
27	Correlación entre características, CI e IV.	66
28	Relación entre IMC y CI.	67

Índice de tablas

1	Estadísticos resumen del peso, la altura y la edad de los participantes.	31
2	Estadísticos resumen para PSQI.	34
3	Estadísticos resumen para DSI.	36
4	Parámetros del modelo FMM para los usuarios.	42
5	Métricas de silueta para diferentes métodos de clustering con los parámetros del modelo FMM.	43
6	Parámetros del modelo FMM para los usuarios del cluster 2 obtenido con método jerárquico <i>single</i>	44
7	Parámetros del modelo FMM para los usuarios y clusters de K-medias.	46
8	Métricas de silueta para diferentes métodos de clustering con la distancia DTW y ACPF.	54
9	Usuarios en clusters 1 y 2 usando K-medias, DTW y ACPF.	55
10	Índices no paramétricos para la Frecuencia Cardíaca (1).	59
11	Índices no paramétricos para la Frecuencia Cardíaca (2).	60
12	Valores de IV para cada usuario ordenados de mayor a menor.	62
13	Valores de CI para cada usuario ordenados de mayor a menor.	64
14	Estadísticos resumen de las variables del archivo sleep.csv.	72
15	Estadísticos resumen del IBI para cada usuario.	73
16	Estadísticos resumen para el cuestionario MEQ.	73
17	Clasificación de usuarios basada en las puntuaciones de MEQ.	74
18	Clasificación de usuarios basada en las puntuaciones de Pittsburgh.	74
19	Estadísticos resumen de la frecuencia cardíaca para cada usuario.	75

1 Introducción

La alteración del ritmo circadiano en los seres humanos se ha vinculado con la aparición de enfermedades crónicas y el empeoramiento de condiciones preexistentes como la diabetes, la hipertensión o la depresión [1–3]. Por esta razón, en las últimas décadas, se ha incrementado la investigación sobre el sistema circadiano humano. El desarrollo de la tecnología ha permitido desarrollar técnicas no invasivas que permiten recopilar información continua de biomarcadores o variables de interés en la biología circadiana, durante las 24 horas del día e incluso durante varios días consecutivos.

Actualmente, dispositivos portátiles como relojes inteligentes y bandas de pecho registran datos sobre la actividad física, la presión arterial y la frecuencia cardíaca, entre otros. Estos datos exhiben patrones oscilatorios a lo largo del día y están sujetos a variabilidad y a interrupciones, como las causadas por la descarga de la batería del dispositivo o la retirada momentánea del dispositivo. Esos dispositivos son capaces de indicar si se ha sufrido una caída, qué ejercicio hemos realizado y avisarnos de un posible problema de salud.

En la actualidad existen multitud de bases de datos que recogen valiosa información sobre los ritmos diarios de los individuos. Uno de los conjuntos más destacados en este contexto es UK Biobank, que contiene datos de más de 100,000 participantes, proporcionando información detallada sobre la actividad física, los patrones de sueño y otros marcadores de salud. Este conjunto tiene una complejidad elevada, pues integra una vasta cantidad de datos clínicos, genéticos e imágenes. En esta línea, otro conjunto de interés es MMASH (*Multilevel Monitoring of Activity and Sleep in Healthy People*) [4], que recopila información sobre variables circadianas a lo largo de 24 horas para 22 participantes. Aunque MMASH es menos complejo, su tamaño abordable y enfoque específico permiten un análisis más controlado y detallado de los ritmos circadianos, combinando datos objetivos, como la frecuencia cardíaca y biomarcadores, con datos subjetivos obtenidos a través de diversos cuestionarios.

La disponibilidad de estos conjuntos de datos ha impulsado la investigación en cronobiología y medicina circadiana, permitiendo a los científicos identificar correlaciones entre los ritmos circadianos y diversas condiciones de salud, y desarrollar intervenciones personalizadas para mejorar el bienestar general. Estos avances han sido fundamentales para mejorar nuestra comprensión de cómo los ritmos circadianos afectan la salud y el comportamiento, y han facilitado la implementación de intervenciones basadas en cronoterapia para tratar trastornos del sueño, cáncer, y otras condiciones relacionadas con la disrupción circadiana [5].

Desde el punto de vista del análisis de estos datos, tradicionalmente se han utilizado los métodos no paramétricos debido a su simplicidad, a pesar de ser más susceptibles al ruido o valores extremos. Existe una variedad de índices no paramétricos clásicos, que se calculan directamente a partir de los datos, y se utilizan rutinariamente en la medicina circadiana para monitorizar y analizar la actividad diaria y el sueño [6]. Entre los índices más empleados se encuentran la estabilidad inter-diaria (IS), que mide la consistencia del ritmo a lo largo de varios días; la variabilidad intra-diaria (IV), que evalúa la fragmentación del ritmo; el valor promedio y el momento de las cinco horas consecutivas con los valores más bajos (VL5 y L5); el valor promedio y el momento de las diez horas consecutivas con los valores más altos (VM10 y M10); y la amplitud relativa (RA), que es la diferencia entre VM10 y VL5 dividida por $VM10+VL5$.

En esta misma línea, uno de los índices o medida de mayor trascendencia en la investigación más reciente es el denominado TAP (Temperatura, Actividad y Posición), que proporciona una evaluación integral de los ritmos circadianos al combinar datos de temperatura corporal, actividad física y posición corporal. La eficiencia de esta medida se ha demostrado en estudios de polisomnografía [7] o en la evaluación de cronotipos [8]. Sin embargo, no siempre se dispone de los datos suficientes para poder calcularla.

Además de los anteriores, también se han explorado otros índices más elaborados basados en el enfoque del análisis espectral [9], [10], así como métodos no paramétricos de análisis de series temporales adaptativas [11], proporcionando herramientas robustas para monitorizar la actividad diaria y el sueño, pero resultando a menudo de difícil interpretación en la práctica clínica. Todos estos índices caracterizan únicamente aspectos parciales de los ritmos biológicos de interés, en ocasiones carecen de interpretación fisiológica, y están sujetos al tipo de datos, frecuencia de muestreo, complejidad de los datos, ruido y preprocesado, limitando o cuestionando su uso de forma generalizada.

Desde la perspectiva de la modelización, las ecuaciones diferenciales han sido utilizadas para capturar dinámicas no lineales, permitiendo modelar interacciones entre múltiples ritmos biológicos, así como su variabilidad a lo largo del tiempo [12]. Sin embargo, estos modelos dependen de parámetros desconocidos que necesitan ser calibrados, lo cual generalmente se realiza estimando dichos parámetros a partir de datos experimentales. Esto hace que, en muchas ocasiones, las predicciones no sean tan precisas como se desearía, además de resultar de difícil interpretación. En los últimos años, ha proliferado el uso de modelos paramétricos, los cuales asumen una distri-

bución sobre la forma de los datos y están definidos por un conjunto de parámetros fijos y finitos que describen completamente la distribución de la variable de interés. Estos modelos presentan una estructura clásica de señal más ruido, y se han extendido en la práctica debido a su flexibilidad y capacidad para capturar y describir de manera sencilla los principales aspectos de los ritmos biológicos. El modelo más ampliamente utilizado en este campo es el modelo cosinor [13], que ajusta los datos a una señal sinusoidal. El modelo cosinor se caracteriza por tres parámetros y su principal limitación radica en que no proporciona resultados adecuados cuando los datos presentan asimetría. Frente a este inconveniente, en los últimos años, el grupo de investigación “Inferencia con Restricciones” de la Universidad de Valladolid ha desarrollado y publicado diversas extensiones del modelo estadístico *Frequency Modulated Möbius* (FMM) para el análisis de señales oscilatorias [14, 15]. Este modelo paramétrico es flexible y capaz de capturar, a partir de cinco parámetros, patrones asimétricos y complejos con una o más ondas que habitualmente subyacen en la biología circadiana. Además, los parámetros del modelo FMM tienen una clara interpretación fisiológica. Su eficacia ha sido demostrada en aplicaciones en cardiología y astrología, entre otras [14].

En la práctica, las metodologías de análisis se combinan o adaptan en función del conjunto de datos y las variables de interés. En el caso de MMASH, por ejemplo, la falta de información para calcular el índice TAP obliga a desarrollar métodos alternativos para estudiar variables clave del ritmo circadiano, como la frecuencia cardiaca, los niveles hormonales, el movimiento y los cambios de posición corporal. Estas variables presentan patrones oscilatorios complejos, que a menudo se ven interrumpidos por ruido, periodos de inactividad o datos faltantes. La modelización precisa de estos ritmos, así como su interpretación fisiológica, supone un desafío considerable, ya que los modelos tradicionales, como el cosinor, no son adecuados para datos asimétricos o irregulares. En este contexto, el desarrollo de modelos avanzados como el FMM, junto con el estudio de índices no paramétricos adaptados a la variabilidad biológica, resulta crucial para abordar estas limitaciones y mejorar nuestra comprensión de los ritmos circadianos.

Este trabajo tiene tres objetivos principales: 1) demostrar la eficacia del modelo FMM en el análisis de señales circadianas complejas obtenidas de la actividad física y la frecuencia cardiaca y en base a sus parámetros, clasificar los distintos usuarios; 2) desarrollar una nueva medida y estudiar métodos no paramétricos para analizar la variabilidad de las señales biológicas; y 3) calcular índices no paramétricos y establecer relaciones con la información disponible en el conjunto de datos MMASH.

2 Métodos

En esta sección presentamos los métodos paramétricos y no paramétricos que se emplearan en este trabajo. Por un lado, se presentan dos modelos paramétricos para señales oscilatorias cuyo uso se ha extendido en cronobiología en los últimos años. Por otro, se exponen varios índices no paramétricos de interés en este trabajo, junto con la metodología no paramétrica del Análisis de Componentes Principales Funcionales (ACPF). Por último, se desarrolla una nueva medida para el análisis de datos circadianos.

2.1 Modelos paramétricos

Sea $X(t)$ el valor de expresión de un determinado ritmo biológico (frecuencia cardiaca, actividad corporal, etc.) a lo largo del tiempo, donde $t \in [0, 2\pi]$. En caso de que el tiempo t' se encuentre en otro intervalo $[t_0, T + T_0]$, podemos transformarlo mediante la siguiente expresión: $\frac{(t'-t_0)}{T} \cdot 2\pi$ donde t_0 es el valor inicial del intervalo y T representa la duración total del mismo. Los modelos para señales oscilatorias se definen como un modelo clásico de señal más ruido:

$$X(t) = \mu(t) + e(t)$$

donde $\mu(t)$ es una señal oscilatoria (ver Definición 1) y $e(t)$ representa el término del error en el tiempo t .

Definición 1 (Señal Oscilatoria). Una señal oscilatoria es la medida de la trayectoria de una partícula en un sistema oscilatorio, en el que la partícula vuelve al estado inicial después de un período. Matemáticamente, una señal oscilatoria en el espacio euclídeo $\mu(t)$ se puede formular de manera equivalente como una señal $\varphi(t)$ en el círculo unidad (espacio circular) de la siguiente forma [16]:

$$\mu(t) = \cos(\varphi(t)), \quad 0 < \varphi(t) \leq 2\pi, \quad 0 < t \leq 2\pi$$

Así, $X(t)$ representa los datos observados a lo largo del tiempo para los que subyace una señal oscilatoria $\mu(t)$.

A continuación, se describe el modelo cosinor, y el modelo FMM para el análisis de señales oscilatorias.

2.1.1 Modelo cosinor

El modelo cosinor fue introducido inicialmente en [17], mediante una formulación matemática basada en ondas sinusoidales. La expresión matemática del modelo es la siguiente:

$$X(t) = \mu(t) + e(t) = M + A \cos(t + \phi) + e(t), \quad t \in [0, 2\pi]$$

donde M representa el valor medio de la onda, conocido como *Midline Statistic Of Rhythm* (MESOR), A denota la amplitud, ϕ es la acrofase, es decir, el tiempo transcurrido desde el comienzo de la onda hasta el punto máximo y $e(t)$ es el término del error.

La principal limitación de este modelo es que solo puede describir ondas simétricas. Aunque el modelo tiene diversas extensiones que permiten el ajuste de múltiples componentes sinusoidales o la incorporación de covariables, tales adaptaciones no se consideran en el presente trabajo.

2.1.2 Modelo FMM

El modelo FMM fue introducido inicialmente en [14], mediante una formulación matemática que incluye parámetros para la forma de la señal, lo que permite describir una mayor variedad de patrones, más allá de ondas simétricas. La expresión matemática del modelo es la siguiente:

$$X(t) = \mu(t) + e(t) = M + A \cos(\phi(t)) + e(t), \quad t \in [0, 2\pi]$$

donde:

- $M \in \mathbb{R}$, $A \in \mathbb{R}^+$.
- $\phi(t) = \beta + 2 \arctan \left(\omega \tan \left(\frac{t-\alpha}{2} \right) \right)$; $\alpha, \beta \in [0, 2\pi]$.
- $\omega \in [0, 1]$.
- $e(t) \sim \mathcal{N}(0, \sigma^2)$.

Los distintos aspectos de la onda son descritos por estos cinco parámetros:

- M : es el intercepto del modelo que indica el nivel basal de la onda.
- A : representa la amplitud de la onda.

- α : este parámetro controla el inicio de la onda dentro de su ciclo.
- β y ω : son parámetros de la forma de la onda. β representa la simetría de la onda, cuando β es un múltiplo entero de π , la onda es simétrica. Además, señala si la onda presenta una estructura ascendente-descendente o descendente-ascendente en los momentos de máxima amplitud. Por otro lado, ω indica el apuntamiento de la onda; $\omega = 0$ indica una onda muy apuntada, mientras que $\omega = 1$ indica una onda sinusoidal.

La flexibilidad del modelo FMM se debe al término $\phi(t)$. Esta variable angular captura el movimiento periódico de la oscilación mediante tres parámetros lineales, de los cuales dos son angulares, a diferencia del modelo cosinor, en el cual este corresponde a una sola variable lineal.

Otros parámetros importantes que se pueden obtener del modelo FMM son los instantes en los que el modelo alcanza sus valores máximos y mínimos de expresión, denotados por t_U y t_L , respectivamente:

$$t_U = \alpha + 2 \arctan \left(\frac{1}{\omega} \tan \left(\frac{-\beta}{2} \right) \right)$$

$$t_L = \alpha + 2 \arctan \left(\frac{1}{\omega} \tan \left(\frac{\pi - \beta}{2} \right) \right)$$

Adecuación del modelo

Para evaluar la calidad del ajuste del modelo, se utiliza el coeficiente de determinación R^2 . Este coeficiente varía entre 0 y 1, indicando un mejor ajuste cuanto más cercano esté a 1. El R^2 para una onda se define como:

$$R^2 = 1 - \frac{\sum_t (X(t) - \hat{\mu}(t))^2}{\sum_t (X(t) - \bar{X})^2}$$

donde $\hat{\mu}(t)$ es la señal oscilatoria predicha y \bar{X} es el valor medio de la oscilación.

2.2 Métodos no paramétricos

En esta sección, se presentan los índices no paramétricos que se emplearan en este trabajo, así como el enfoque del Análisis de Componentes Principales Funcionales (ACPF).

2.2.1 Índices no paramétricos

Los índices no paramétricos son frecuentemente empleados para describir los ritmos circadianos, particularmente en investigaciones que involucran a seres humanos, como los presentados en [18] y [19]. A pesar de que estos índices son relativamente fáciles de implementar y no requieren un alto nivel de experiencia por parte del usuario, en algunos casos estas medidas pueden carecer de rigor matemático y resultar ineficaces, por ejemplo, cuando existen múltiples oscilaciones o el nivel de ruido es elevado.

En el análisis de datos circadianos, generalmente los índices se agrupan en cinco categorías

- Índices relacionados con la amplitud del ritmo.
- Índices relacionados con la temporalidad.
- Índices que describen la regularidad del ritmo.
- Índices que indican la sincronización del ritmo con el entorno.
- Índices que caracterizan la robustez del sistema circadiano.

El lector puede encontrar una descripción más detallada de estos y otros índices en [19].

Amplitud del Ritmo

La amplitud de un ritmo circadiano describe la variación en los valores de una variable entre los momentos de máxima y mínima actividad en un ciclo de 24 horas. La amplitud se ha asociado con la estabilidad del ritmo, dado que una amplitud reducida generalmente sugiere un ritmo menos robusto.

La amplitud de una variable circadiana se calcula utilizando diferentes fórmulas para los períodos de mayor y menor actividad, debido a las diferencias en la duración y la intensidad de estos períodos. Los períodos de mayor actividad, que generalmente coinciden con las horas de vigilia, son aquellos en los que la actividad fisiológica y metabólica es más intensa. En contraste, los períodos de menor actividad corresponden a las horas de descanso o sueño, durante los cuales el organismo se encuentra en un estado más pasivo. Por ejemplo, la frecuencia cardíaca disminuye durante las horas de menor actividad, mientras que la temperatura de la piel suele aumentar en este mismo período. Por lo tanto, se utilizan fórmulas específicas para representar con

precisión las variaciones entre los niveles de actividad en estos diferentes momentos del ciclo circadiano.

Los siguientes índices se utilizan para calcular la amplitud:

- **VM5:** Promedio de la variable durante las 5 horas consecutivas con los valores más altos.
- **VL5:** Promedio de la variable durante las 5 horas consecutivas con los valores más bajos.
- **VM10:** Promedio de la variable durante las 10 horas consecutivas con los valores más altos.
- **VL10:** Promedio de la variable durante las 10 horas consecutivas con los valores más bajos.

A continuación, se utilizan las variables anteriores para definir la amplitud relativa y absoluta.

- **Amplitud absoluta.** El cálculo de la amplitud varía según el tipo de variable estudiada:
 - VM5 - VL10, para aquellas variables que alcanzan sus valores máximos durante el período de menor actividad.
 - VM10 - VL5, para aquellas variables que alcanzan sus valores máximos durante el período de mayor actividad.
- **Amplitud relativa (RA).** Se calcula de la siguiente manera, según el tipo de variable:

- Para aquellas variables cuyos valores aumentan durante el período de menor actividad:

$$RA = \frac{VM5 - VL10}{VM5 + VL10}$$

- Para aquellas variables cuyos valores aumentan durante el período de mayor actividad:

$$RA = \frac{VM10 - VL5}{VM10 + VL5}$$

El valor de RA está acotado por 1, lo que ocurre solo cuando VL5 o VL10 es 0. Aunque es una medida popular, presenta ciertas inexactitudes, especialmente cuando VL10 o VL5 tienen valores elevados, lo que afecta la precisión de RA. Además, RA depende considerablemente de los valores mínimos y máximos de las variables. En el caso de variables como la actividad motora o la exposición a la luz, cuyos valores durante el sueño son casi nulos, RA no discrimina adecuadamente entre individuos con diferentes niveles de actividad durante la vigilia. Para mitigar estas limitaciones, se ha propuesto la amplitud relativa normalizada (NRA) [19] como una alternativa, utilizando los percentiles (P95) y (P5) de cada variable.

Temporalidad del Ritmo

Los índices que se presentan a continuación se utilizan para determinar la posición de un ritmo dentro de un marco temporal. Existen marcadores de fase tanto para los períodos de menor y mayor actividad. Los más empleados son:

- **TM5**: Tiempo medio de las 5 horas consecutivas con los valores de la variable más altos.
- **TL5**: Tiempo medio de las 5 horas consecutivas con los valores de la variable más bajos.
- **TM10**: Tiempo medio de las 10 horas consecutivas con los valores de la variable más altos.
- **TL10**: Tiempo medio de las 10 horas consecutivas con los valores de la variable más bajos.

Sin embargo, estos índices pueden tener limitaciones, como la sensibilidad a la variabilidad individual y la posible falta de representatividad en casos con patrones de actividad atípicos. Estas restricciones pueden afectar la precisión en la localización del ritmo circadiano.

Regularidad y Fragmentación del Ritmo Circadiano

La regularidad de un ritmo circadiano a lo largo del tiempo es una medida que permite evaluar su estabilidad.

- **Variabilidad intradiaria (IV)**: Mide la fragmentación del ritmo, es decir, la frecuencia de transiciones entre los periodos de mayor y menor actividad. Se

calcula como la división entre la media cuadrática de la diferencia entre todas las horas sucesivas y la media cuadrática alrededor de la media general:

$$IV = \frac{n \sum_{i=2}^n (X(t_i) - X(t_{i-1}))^2}{(n-1) \sum_{i=1}^n (X(t_i) - \bar{X})^2}$$

donde $X(t_i)$ denota el valor de la variable en el instante de tiempo i -ésimo, $1 \leq i \leq n$. Los valores de IV son cercanos a 0 para ondas sin fragmentación (sinusoidales o no) y cercanos a 2 para la máxima fragmentación. Sin embargo, IV es muy sensible al intervalo de muestreo, siendo necesario ajustar la frecuencia del muestreo cuando se calculan IV entre diferentes dispositivos.

- **Índice de fragmentación de sueño (IFs):** Este índice se calcula como el número de despertares dividido por el tiempo total de sueño. Es especialmente útil para evaluar la calidad del sueño, ya que ayuda a determinar si un sueño es continuo o se interrumpe frecuentemente.

$$IFs = \frac{\text{Número de despertares}}{\text{Duración total de sueño}}$$

Un mayor valor de IFs indica una mayor fragmentación del sueño, lo que generalmente se asocia con una peor calidad de sueño y un menor descanso. Un inconveniente del índice de fragmentación del sueño es que es muy sensible a pequeños despertares o movimientos durante la noche, incluso si estos no resultan en una interrupción significativa del descanso.

Existen otros índices, como la Estabilidad Interdiaria (IS), que miden la robustez del ritmo a lo largo de los días, ver [19, 20].

Indicadores Compuestos para Evaluar la Salud y la Solidez del Sistema Circadiano

A continuación, explicamos una métrica para analizar la robustez del sistema circadiano.

- **Índice de Circadianidad (CI)** Este índice se utiliza para cuantificar la robustez del ritmo circadiano de una variable específica. La expresión matemática de CI es la siguiente:

$$CI = \frac{P_1}{\sum_{i=1}^{12} P_i}$$

Se define como la proporción de contribución del primer armónico, que tiene un ciclo de 24 horas, en comparación con la suma de los primeros 12 armónicos, calculados mediante un análisis de Fourier.

Un valor de CI cercano a 1 indica un sistema circadiano bien desarrollado y estable, donde la mayor parte de la variabilidad en la señal es consecuencia de un ritmo circadiano de 24 horas. Por el contrario, un CI bajo indica que otros ciclos, menores a 24 horas, tienen un impacto considerable en la señal, resultando en una mayor variabilidad no alineada con el ritmo circadiano estándar.

Existen otras métricas, como el Índice de Función Circadiana (CFI) y el Índice de Salud Circadiana (CHI), que también son ampliamente utilizadas para evaluar la solidez del sistema circadiano. Para más detalles y aplicaciones, ver [19, 20].

2.2.2 Análisis de Componentes Principales Funcionales

El análisis de señales oscilatorias se puede abordar desde la perspectiva del análisis funcional, donde cada dato se define como una función. Así, $X(t)$ representa los datos observados a lo largo del intervalo $[0, 2\pi]$ para los cuales subyace una señal oscilatoria $\mu(t)$.

Esta expresión también se puede tratar como un dato funcional, es decir, aquellos datos en los que cada observación se define a partir de una función. Por lo tanto, cada dato funcional para señales oscilatorias está definido por $X(t)$, la señal oscilatoria y los diferentes instantes t en los que ocurre.

El *Análisis de Componentes Principales Funcionales* (ACPF) es una extensión del análisis de componentes principales (ACP) tradicional que se utiliza cuando los datos son funciones, curvas o series temporales. Este tipo de análisis es particularmente útil en situaciones donde las observaciones están definidas a lo largo de una variable continua, como puede ser el tiempo o el espacio, y se representan mejor mediante funciones que capturan la evolución o variación de los datos a lo largo de dicha variable. Una introducción al análisis de datos funcionales se puede ver en [21].

El ACPF es considerado un método no paramétrico, ya que no asume una forma paramétrica específica para las funciones subyacentes. En lugar de ello, utiliza bases funcionales, como splines o funciones de Fourier, para aproximar las funciones observadas, proporcionando flexibilidad en la representación de los datos sin imponer una estructura rígida.

ACPF busca reducir la dimensionalidad del espacio funcional mediante la identificación de componentes principales que capturen la mayor parte de la variabilidad de los datos. Estas componentes principales son funciones ortogonales que maximizan la varianza explicada en el conjunto de datos, permitiendo una representación eficiente y compacta de la información.

En el apéndice A se puede encontrar una introducción matemática, mencionando los teoremas necesarios y definiciones para explicar el ACPF.

Este enfoque considera las muestras de datos funcionales como funciones aleatorias. La situación es la siguiente: sea n el número de trayectorias observadas, cada una representada por $X_i(t)$ para $1 \leq i \leq n$ y $t \in [0, T]$, donde T es el límite superior del intervalo de tiempo de observación.

Cada $X_i(t)$ es una realización de un proceso estocástico $X(t)$, que modela la evolución temporal de una variable aleatoria a lo largo del tiempo. Un proceso estocástico es una colección de variables aleatorias, una para cada instante t , que describe la evolución aleatoria de un fenómeno a lo largo del tiempo. Por lo tanto, en este enfoque $X(t)$ no es una única función, sino una familia de funciones que varían con el tiempo.

Las trayectorias individuales $X_i(t)$ son realizaciones del proceso estocástico subyacente $X(t)$, observadas para cada usuario. Estas trayectorias pueden exhibir diferentes comportamientos dependiendo de la variabilidad inherente al proceso estocástico. Para analizar este conjunto de trayectorias, se utiliza una base funcional formada por las autofunciones del operador de covarianza del proceso $X(t)$, que nos permite representar y estudiar las trayectorias en un espacio funcional adecuado.

La función de covarianza $K_X(s, t)$ de un proceso estocástico $X(t)$ describe la dependencia conjunta entre los valores del proceso en dos instantes diferentes s y t . Formalmente, se define como:

$$K_X(s, t) = \text{Cov}(X(s), X(t)) = \mathbb{E}[(X(s) - \mathbb{E}[X(s)])(X(t) - \mathbb{E}[X(t)])],$$

donde $\mathbb{E}[X(t)]$ es la función media del proceso.

Para aplicar los siguientes teoremas, se asume que el proceso estocástico está centrado. En este contexto, se considera $X_i(t) - \rho(t)$ en lugar de $X_i(t)$, donde $\rho(t)$ es la función media del proceso estocástico $\mathbb{E}[X(t)]$.

Se considera el operador de covarianza $T_{K_X} : L^2[0, T] \rightarrow L^2[0, T]$ que es un operador compacto en un espacio de Hilbert.

Por el Teorema de Mercer 2, el núcleo de T_{K_X} , es decir, la función de covarianza $K_X(\cdot, \cdot)$, tiene una descomposición espectral:

$$K_X(s, t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t),$$

donde la serie converge absolutamente y uniformemente, y λ_k son valores propios no negativos ordenados de mayor a menor con sus correspondientes autofunciones ortonormales $e_k(t)$, es decir, tienen norma 1 (recordamos que estamos en L^2) y $\int_0^T e_k(t) e_l(t) dt = 0$. Por el Teorema de Karhunen–Loève 3, la expansión ACPF de una trayectoria aleatoria subyacente es:

$$X_i(t) = \rho(t) + \sum_{k=1}^{\infty} Z_{ik} e_k(t),$$

donde

$$Z_{ik} = \int_0^T (X_i(t) - \rho(t)) e_k(t) dt$$

son los scores o puntuaciones.

La expansión de Karhunen–Loève facilita la reducción de dimensionalidad en el sentido de que la suma parcial converge uniformemente, es decir,

$$\sup_{t \in [0,1]} \mathbb{E} \left[X_i(t) - \rho(t) - \sum_{k=1}^K Z_{ik} e_k(t) \right]^2 \rightarrow 0$$

cuando que $K \rightarrow \infty$, y por lo tanto, la suma parcial con un K suficientemente grande proporciona una buena aproximación a la suma infinita. De esta manera, la información en X_i se reduce de dimensión infinita a un vector de dimensión K :

$$Z_i = (Z_{i1}, \dots, Z_{iK}),$$

con el proceso aproximado:

$$X_i^{(K)}(t) = \rho(t) + \sum_{k=1}^K Z_{ik} e_k(t) \tag{1}$$

La varianza explicada por cada una de las autofunciones se calcula igual que en el caso tradicional, dividiendo el autovalor correspondiente entre la suma del total. Para elegir las autofunciones ortonormales $e_k(t)$, es común utilizar bases spline, series de Fourier y bases wavelet.

ACPF puede ser utilizado en caso de que los datos sean discretos e irregulares para cada una de las trayectorias. En nuestro caso analizaremos las autofunciones y estudiaremos el proceso aproximado.

2.2.3 Desarrollo de una nueva medida

A continuación, se presenta una medida no paramétrica construida a partir de tres variables clave que reflejan la actividad circadiana diaria: el movimiento, registrado desde la muñeca, la frecuencia cardiaca y los cambios en la posición corporal.

El movimiento refleja la actividad física diaria, proporcionando información sobre los períodos de mayor y menor actividad, así como sobre la cantidad total de actividad realizada a lo largo del día. La frecuencia cardiaca, estrechamente vinculada al sistema nervioso autónomo, sigue un patrón circadiano, con valores más bajos durante los períodos de sueño y más altos durante los períodos de vigilia. Por su parte, los cambios en la posición corporal, como estar de pie, sentado o acostado, permiten identificar momentos de descanso o actividad y facilitan la distinción entre las distintas fases del ciclo diario, lo que resulta crucial para comprender cómo el cuerpo responde a la regulación circadiana. Denotaremos el movimiento como “Mv”, la frecuencia cardiaca como “FC” y los cambios de posición como “w”.

La construcción de esta medida está inspirada por la medida TAP [20], que combina la temperatura, la actividad y la posición de la muñeca en el espacio. En este caso, la actividad motora se cuantifica a través del cambio en grados de la posición, lo que permite evaluar la movilidad a partir de variaciones angulares. La temperatura refleja las fluctuaciones térmicas a lo largo del día, mientras que la posición de la muñeca se determina mediante el ángulo de inclinación respecto al plano horizontal (ver [22]).

En ausencia de estas variables, hemos desarrollado una medida alternativa, wMvFC, que permite evaluar los ritmos circadianos minimizando la influencia de factores externos. Esta nueva medida incluye la frecuencia cardiaca, medida en latidos por minuto, el movimiento de la muñeca, expresado como la magnitud del vector de aceleración en el espacio medida en m/s^2 , y el número de cambios de posición corporal.

Esta medida puede calcularse para cualquier intervalo de tiempo y permite la comparación de valores entre distintos usuarios, siempre que los datos de todos ellos correspondan al mismo intervalo temporal. Sin embargo, si se desea analizar la variación entre los periodos de menor y mayor actividad, es recomendable utilizar un periodo de 24 horas para capturar adecuadamente los ritmos circadianos.

A continuación explicamos como calcular esta medida:

1. Seleccionar los percentiles P5 y P95 del movimiento y de la frecuencia cardiaca para eliminar valores atípicos.
2. Calcular el promedio del movimiento, la frecuencia cardiaca y el número de cambios de posición en intervalos de 15 minutos.
3. Si el número de cambios de posición es igual a 0, sustituir este valor por 0.5, con el objetivo de no anular la medida durante los periodos de menor actividad (como el sueño).
4. Normalizar las variables de movimiento y frecuencia cardiaca usando el método de normalización min-max para que ambas estén en el rango de $[0, 1]$, permitiendo así una combinación adecuada de estas variables.
5. Usar la siguiente fórmula:

$$wMvFC = w \left(\frac{\text{Movimiento} + \text{Frecuencia Cardiaca}}{2} \right)$$

Como resultado, se obtienen tantos valores de $wMvFC$ como intervalos de 15 minutos tengamos disponibles, generando una serie temporal con valores positivos.

La Figura 1 muestra los gráficos de las series temporales de las variables utilizadas en la medida $wMvFC$ para el usuario 3 de los 22 usuarios del conjunto de datos MMASH. El eje X indica las horas en las que se registraron las observaciones.

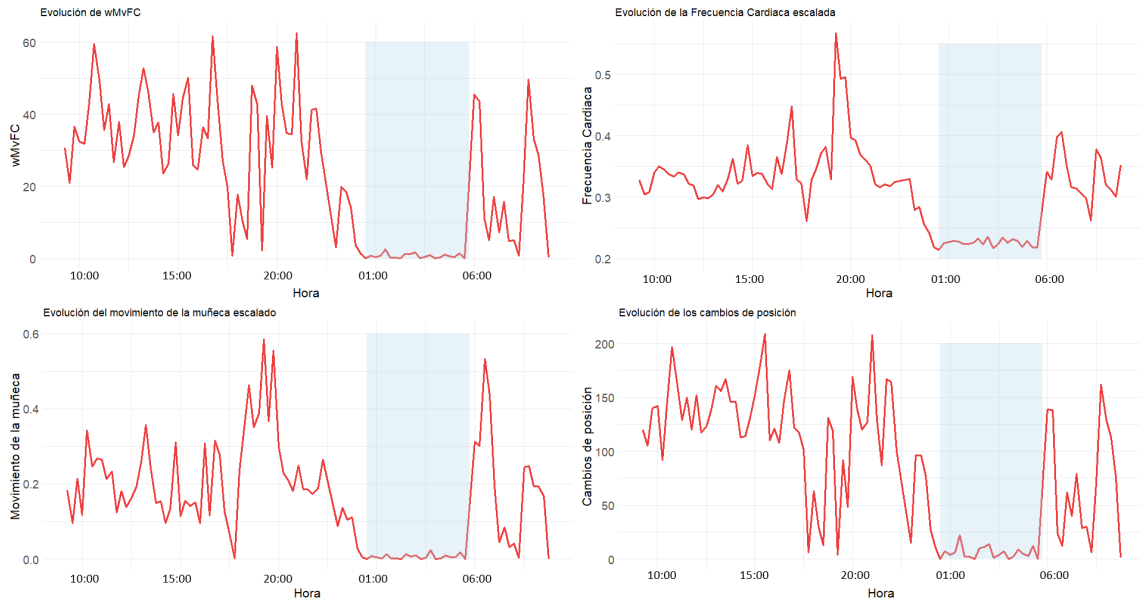


Figura 1: Evolución de las variables de wMvFC para el usuario 3 del conjunto MMASH.

Las horas de sueño del usuario se encuentran resaltadas en azul, correspondiendo a los periodos en los que se alcanzan los valores más bajos.

3 Descripción del conjunto MMASH

A continuación, se describe el conjunto de datos MMASH, el cual es de acceso público bajo la Licencia de Base de Datos Abierta (ODbL) v1.0 y está disponible en PhysioNet: <https://physionet.org/content/mmash/1.0.0/>. La descripción detallada de los datos se basa en [23]. La lectura y procesamiento de los datos se ha realizado utilizando el software estadístico R. Para más detalles sobre el código empleado, consulte el repositorio disponible en <https://github.com/MIGUELCUBEROGOMEZ/TFM>.

3.1 Descripción del conjunto de datos

El conjunto de datos MMASH proporciona información relevante para la biología circadiana, recopilada de 22 participantes y organizada en siete archivos para cada uno de ellos. Los datos abarcan un período ligeramente superior a 24 horas, desde la primera hora de la mañana del día 1 hasta la primera hora de la mañana del día 2, donde “1” representa el primer día y “2” el segundo.

Los archivos son:

- `user_info.csv`: características antropomórficas del participante.
- `sleep.csv`: duración y calidad del sueño del participante.
- `RR.csv`: datos de latido a latido.
- `questionnaire.csv`: puntuaciones de todos los cuestionarios.
- `Activity.csv`: diario de actividades a lo largo del día.
- `Actigraph.csv`: datos del acelerómetro triaxial de muñeca.
- `saliva.csv`: concentraciones de cortisol y melatonina en saliva registradas antes de dormir y justo después de despertar.

A continuación se describe el contenido de cada archivo.

3.1.1 Datos de los usuarios (`user_info.csv`)

- **Género**: masculino (M) o femenino (F).
- **Altura**: altura en centímetros de cada participante.
- **Peso**: peso en kilogramos de cada participante.
- **Edad**: edad en años de cada participante.

Este archivo recoge las características fisiológicas de 22 hombres adultos. La Tabla 1 recoge los estadísticos resumen de dichas variables.

Variable	mean	sd	min	p25	p50	p75	max
Peso	75.0	12.8	60	67	70	80	115
Altura	180.0	8.22	169	175	180	183	205
Edad	26.0	7.12	20	25	27	27.8	40

Tabla 1: Estadísticos resumen del peso, la altura y la edad de los participantes.

Donde “mean” es la media, “sd” es la desviación estándar, “min” es el valor mínimo, “p25” es el percentil 25, “p50” es el percentil 50 (mediana), “p75” es el percentil 75 y “max” es el valor máximo.

3.1.2 Datos con respecto al sueño (sleep.csv)

- `In Bed Date`: día en que el usuario se fue a la cama, 1 o 2.
- `In Bed Time`: hora del día en que el usuario se fue a la cama.
- `Out Bed Date`: día en que el usuario se levantó de la cama, 1 o 2.
- `Out Bed Time`: hora del día en que el usuario se levantó de la cama.
- `Onset Date`: día en que el usuario se quedó dormido, 1 o 2.
- `Onset Time`: hora del día en que el usuario se quedó dormido.
- `Latency`: tiempo en minutos necesario para quedarse dormido.
- `Total Minutes in Bed (TMB)`: minutos totales en la cama.
- `Total Sleep Time (TST)`: minutos totales de sueño.
- `Efficiency`: división entre el tiempo total de sueño y los minutos totales en la cama (TST/TMB).
- `Wake After Sleep Onset (WASO)`: tiempo despierto después de quedarse dormido por primera vez.
- `Number of Awakenings`: número de despertares.
- `Average Awakening Length`: duración promedio de los despertares en minutos.
- `Movement Index (MI)`: número de minutos sin movimiento durante el sueño, expresado como un porcentaje del total del tiempo en fase de sueño.
- `Fragmentation Index (IFs)`: número de minutos con movimiento durante el sueño, expresado como un porcentaje del total del tiempo en reposo.
- `Sleep Fragmentation Index (SFI)`: división entre los índices de Movimiento y Fragmentación (MI/IFs).

La Tabla 14 del material suplementario recoge los estadísticos resumen para el archivo sleep.

3.1.3 Datos de latido (RR.csv)

- Día: día en el que se recogen los datos, 1 o 2.
- Hora: hora en el que se recoge el latido (hh:mm:ss).
- IBI: es el tiempo medido en segundos entre dos latidos cardíacos consecutivos. Este intervalo refleja el tiempo entre los latidos y es un indicador directo de la frecuencia cardiaca.

En la Tabla 15 del material suplementario se presentan los estadísticos resumen de la variable IBI para cada uno de los usuarios.

3.1.4 Datos del cuestionario (questionnaire.csv)

Cuestionario Mañana-Noche (MEQ)

El MEQ (Morningness-Eveningness Questionnaire), desarrollado por James A. Horne y O. Östberg en 1976, clasifica a las personas según su cronotipo en matutino, vespertino o intermedio [24]. Por ejemplo, las personas matutinas tienden a despertarse temprano y muestran una mayor actividad y productividad durante la mañana. Estas personas suelen sentirse más alerta y enérgicas en las primeras horas del día, experimentando un descenso en su nivel de energía y eficiencia a medida que avanza la tarde y la noche. En términos de temperatura corporal, los individuos matutinos presentan temperaturas más bajas en la madrugada, las cuales comienzan a elevarse progresivamente hacia la mañana, alcanzando su punto máximo justo antes del mediodía. Esta elevación de la temperatura corporal contribuye a un estado de mayor alerta y energía en las primeras horas del día. En contraste, los individuos vespertinos tienden a despertar y alcanzar su temperatura máxima en un momento más tardío del día. Los intermedios no tienen una preferencia marcada por horarios tempranos o tardíos y se adaptan con flexibilidad a diferentes horarios de sueño-vigilia.

El MEQ consta de 19 preguntas numéricas o de opción múltiple (escala de 4–5 puntos) que exploran las preferencias de los participantes para realizar diversas actividades, como la hora a la que prefieren despertarse o comenzar a dormir en comparación con sus horarios reales. La puntuación total varía entre 16 y 86: las puntuaciones menores de 41 indican un cronotipo vespertino, los que oscilan entre 42 y 58 corresponden a un cronotipo intermedio, y los superiores a 59 reflejan un cronotipo matutino.

La Tabla 16 del material suplementario recoge los estadísticos resumen para este cuestionario.

La Tabla 17 del material suplementario recoge la clasificación del cronotipo de los usuarios. En dicha tabla se observa que únicamente 2 usuarios presentan cronotipos vespertinos y 4 tienen cronotipos matutinos, mientras que el resto de los usuarios se clasifica como intermedios.

Inventario de Ansiedad Estado-Rasgo (STAI-Y)

El STAI-Y es una herramienta frecuentemente utilizada para evaluar la ansiedad en términos de estado y rasgo. Su aplicación en entornos clínicos permite la identificación de la ansiedad y la diferenciación de trastornos depresivos.

Índice de calidad de sueño de Pittsburg (PSQI)

El Pittsburgh Sleep Quality Index (PSQI), desarrollado por Daniel J. Buysse en 1989, es una herramienta estandarizada para evaluar la calidad y los patrones del sueño en el último mes. Este índice mide siete áreas: calidad subjetiva del sueño, latencia del sueño, duración del sueño, eficiencia habitual, perturbaciones del sueño, uso de medicación para dormir y disfunción diurna. Cada área se evalúa en una escala de 0 a 3, donde una puntuación de 3 indica mayor dificultad. La puntuación total del PSQI varía de 0 a 21, y un valor inferior a 6 se considera indicativo de una buena calidad del sueño (ver [25]).

En la Tabla 18 del material suplementario se clasifican los usuarios del experimento MMASH en función de su valor del test de Pittsburgh. En dicha tabla se observa que 7 usuarios han reportado mala calidad de sueño la noche realizada el cuestionario.

En la Tabla 2 se presenta el resumen de esta variable.

	mean	sd	min	p25	p50	p75	max
PSQI	5.32	1.99	2	4	5	7	9

Tabla 2: Estadísticos resumen para PSQI.

La media de la variable PSQI es relativamente elevada, aunque aún se clasificaría dentro del rango que indica una buena calidad del sueño. La Figura 2 muestra la frecuencia de cada valor obtenido en el test.

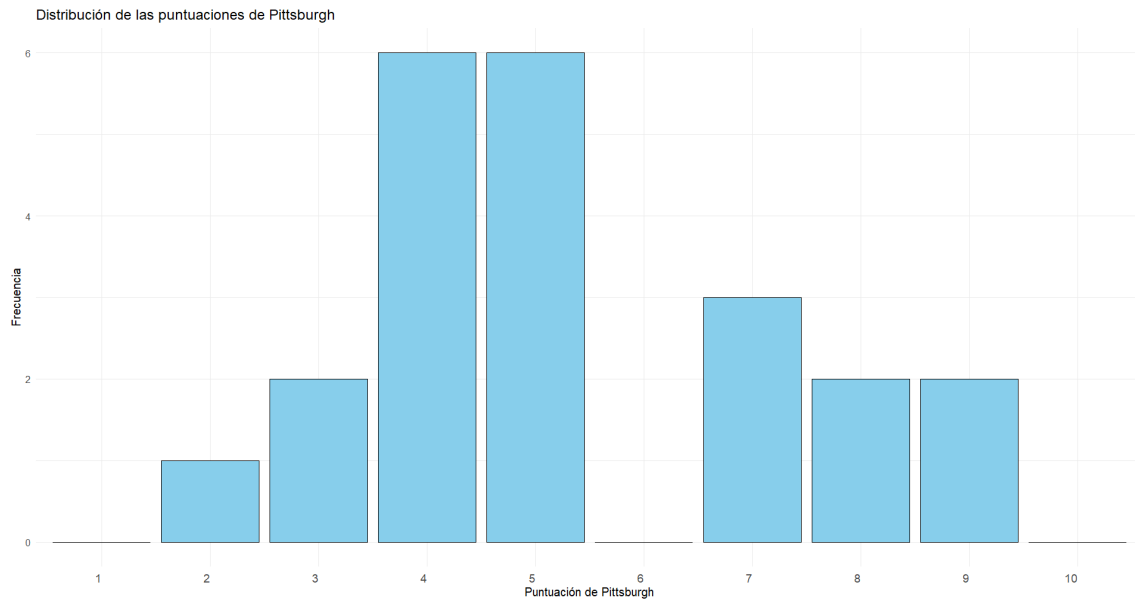


Figura 2: Frecuencia de los valores de PSQI.

En la muestra no se identificaron usuarios con una puntuación de 6, que se sitúa en el umbral entre una buena y una mala calidad del sueño.

Sistemas de Escalas de Inhibición Conductual/Activación Conductual (BIS/BAS)

El cuestionario (BIS/BAS), desarrollado por Carver y White en 1994 estudia la teoría motivacional dual del sistema de activación conductual y el sistema de inhibición conductual. Basado en la teoría de la personalidad de Grey, este cuestionario evalúa cómo los individuos responden a estímulos tanto negativos como atractivos, equilibrando las funciones de activación e inhibición.

Afectaciones positivas y negativas del horario (PANAS)

El afecto positivo y el afecto negativo son aspectos clave que describen la experiencia emocional y el bienestar subjetivo. Ambos son importantes en la investigación y en la práctica clínica para diferenciar ansiedad de depresión y están vinculados a los rasgos de extraversión y neuroticismo, respectivamente.

Inventario diario de estrés (DSI)

El DSI es una escala de autoevaluación compuesta por 58 elementos, diseñada para que los individuos registren eventos estresantes ocurridos en las últimas 24 horas

y evalúen su grado de estrés utilizando una escala de Likert que varía de 1 (“no estresante”) a 7 (“pánico”). La puntuación total se obtiene sumando las calificaciones asignadas a cada elementos.

Las estadísticos resumen de DSI se muestran en la Tabla 3.

	mean	sd	min	p25	p50	p75	max
DSI	32.2	16.3	10	22.2	30.5	40.2	74

Tabla 3: Estadísticos resumen para DSI.

Las puntuaciones son puntuaciones bajas teniendo en cuenta que esta medida tiene un rango entre 1 y 406, observándose en esta muestra un máximo de 74 puntos.

3.1.5 Datos de actividades (activity.csv)

Los participantes registraron todas las actividades realizadas a lo largo del día del experimento en un diario, con horarios de inicio y fin. Las actividades se agrupan en 13 categorías siguiendo esta numeración.

1. Dormir.
2. Estar acostado.
3. Estar sentado (por ejemplo, estudiar, comer y conducir).
4. Movimiento ligero (por ejemplo, caminar lento/medio, tareas domésticas y trabajo).
5. Movimiento medio (por ejemplo, caminar rápido y andar en bicicleta).
6. Movimiento intenso (por ejemplo, gimnasio y correr).
7. Comer.
8. Usar una pantalla pequeña (por ejemplo, teléfono inteligente y computadora).
9. Usar una pantalla grande (por ejemplo, televisión y cine).
10. Consumir bebidas con cafeína (por ejemplo, café o refresco).
11. Fumar.
12. Beber alcohol.
13. Medir los biomarcadores de cortisol y melatonina a través de una muestra de saliva.

La Figura 3 recoge la frecuencia de las actividades a lo largo del experimento.

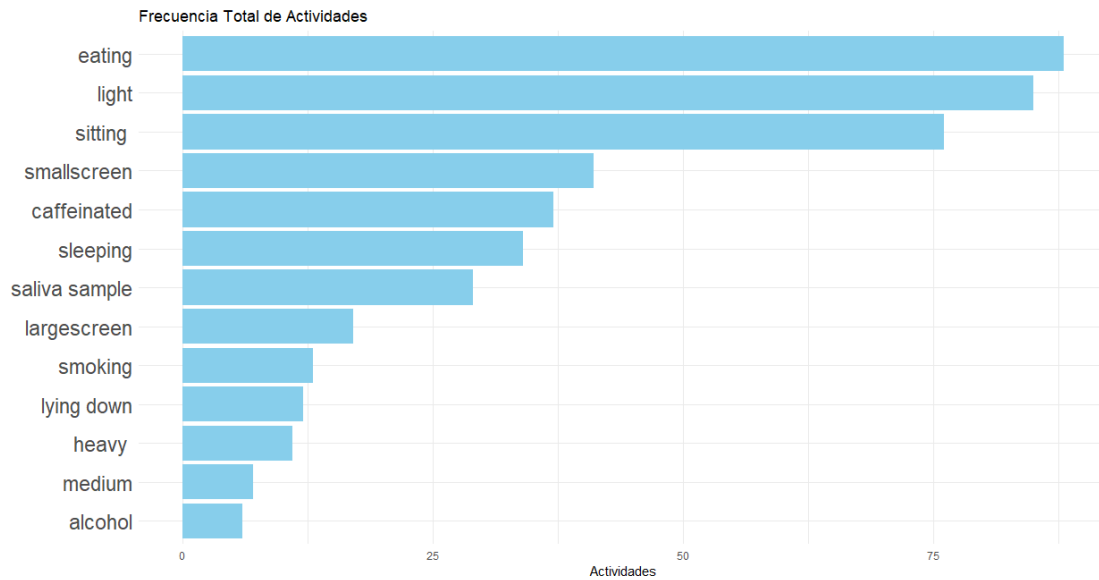


Figura 3: Frecuencia de las actividades.

3.1.6 Datos del acelerómetro (actigraph.csv)

El ActiGraph wGT3X-BT (ActiGraph LLC, Pensacola, FL, EE. UU.) es un acelerómetro triaxial que mide la aceleración en tres ejes perpendiculares: X , Y y Z. Registra la aceleración en cada eje con una frecuencia de muestreo que varía entre 30 y 100 Hz, lo que permite capturar cambios en la aceleración a lo largo del tiempo. Su rango dinámico de ± 8 g indica la capacidad para medir aceleraciones de hasta 8 veces la gravedad terrestre en cualquier dirección. La precisión del dispositivo es de 12 bits, lo que define la resolución de la medición de aceleración, permitiendo detectar variaciones sutiles en la intensidad de la actividad física.

El archivo actigraph.csv contiene las siguientes columnas:

- **Axis1:** aceleración en el eje X expresada en m/s^2 .
- **Axis2:** aceleración en el eje Y expresada en m/s^2 .
- **Axis3:** aceleración en el eje Z expresada en m/s^2 .
- **Pasos:** número de pasos por segundo.
- **Frecuencia cardiaca:** número de latidos cardíacos por segundo.

- **Inclinómetro Apagado:** variable binaria que indica si el inclinómetro está activo (0) o no (1).
- **Inclinómetro de Pie:** variable binaria que muestra si el participante está en posición vertical (1) o no (0) cada segundo.
- **Inclinómetro Sentado:** variable binaria que muestra si el participante está en posición sentada (1) o no (0) cada segundo.
- **Inclinómetro Acostado:** variable binaria que muestra si el participante está en posición acostada (1) o no (0) cada segundo.
- **Magnitud del vector:** magnitud del vector de aceleración, se obtiene como la raíz cuadrada de la suma de los cuadrados de las aceleraciones en los ejes X, Y y Z.
- **día:** día en el que se recogen los datos, 1 o 2.
- **hora:** hora del día en que se registraron los valores de actividad (hh:mm:ss).

La Tabla 19 del amterial suplementario recoge los estadísticos resumen de la frecuencia cardiaca para cada uno de los usuarios.

3.1.7 Datos de las pruebas de saliva (saliva.csv)

Las mediciones de melatonina y cortisol a partir de muestras de saliva se proporcionan en el archivo saliva.csv. Se recogieron dos muestras de saliva por participante. La primera se recogió justo antes de dormir y la segunda justo después de despertarse. Los niveles de melatonina y cortisol están medidos en μg de melatonina por μg de proteínas y en μg de cortisol por 100 μg de proteína, respectivamente.

En las Figuras 4 y 5 se presenta una comparación entre las muestras tomadas por la noche y las recogidas a la mañana siguiente. A la izquierda, en rojo, se muestran los valores correspondientes a las mediciones realizadas antes de dormir, y a la derecha, en azul, los valores tomados recién al despertarse. Cada figura contiene un diagrama de caja que ilustra la distribución de los datos para ambas condiciones.

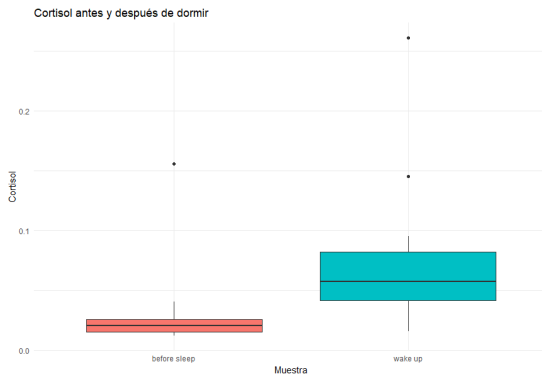


Figura 4: Gráfico de cortisol.

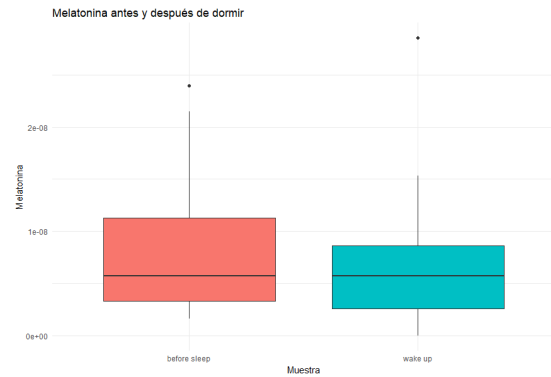


Figura 5: Gráfico de melatonina.

En el gráfico de cortisol, se observa una diferencia notable entre la distribución de las mediciones realizadas antes de conciliar el sueño y aquellas obtenidas al despertar.

4 Resultados

En esta sección se presenta el preprocesamiento de los datos del conjunto MMASH, los resultados obtenidos al aplicar modelos paramétricos a los valores de la medida wMvFC, el análisis mediante componentes principales funcionales para wMvFC, así como los índices no paramétricos correspondientes.

Los resultados se han obtenido utilizando código en R, disponible en el siguiente enlace: <https://github.com/MIGUELCUBEROGOMEZ/TFM>.

En primer lugar, se aplicará el modelo FMM a la medida wMvFC para cada uno de los usuarios. A continuación, se llevará a cabo un análisis de agrupamiento (clustering) basado en los parámetros del modelo FMM, con el fin de clasificar a los usuarios en grupos según los parámetros de la onda ajustada. Posteriormente, se comparará la onda ajustada por dicho modelo con la obtenida mediante el modelo cosinor en dos usuarios seleccionados.

Seguidamente, se empleará el análisis de componentes principales funcionales (ACPF) aplicado a la medida wMvFC. A partir de la aproximación de las series temporales obtenida mediante la Fórmula 1, se realizará un análisis de agrupamiento de las 22 series temporales (una por usuario) con el fin de identificar patrones o tendencias en las series reconstruidas. Posteriormente, se procederá a comparar los resultados obtenidos para dos usuarios específicos con el ajuste proporcionado por el modelo

FMM, evaluando las diferencias entre la aproximación mediante ACPF y el ajuste de dicho modelo.

Finalmente, se calcularán los índices no paramétricos y se analizará su relación con las características individuales de los usuarios y sus respuestas en los cuestionarios.

4.1 Preprocesado

El preprocesado de los datos es una etapa fundamental en cualquier análisis, especialmente cuando se trabaja con datos fisiológicos y de actividad como en el conjunto MMASH. Asegurar la calidad y coherencia de los datos es esencial para obtener resultados fiables y evitar sesgos que puedan distorsionar las conclusiones. En este caso, el preprocesado permite identificar y tratar valores atípicos (outliers), manejar datos faltantes y unificar diferentes fuentes de información.

A continuación, se describe cómo se ha llevado a cabo el preprocesamiento de los datos y la identificación de outliers:

Los datos estaban organizados en siete carpetas diferentes para cada usuario. Como primer paso, se unificaron en tablas conjuntas, una para cada tipo de información contenida en las carpetas. Posteriormente, se realizó una limpieza de los datos en Excel, eliminando registros redundantes, como la variable que indica el número de fila correspondiente.

Dado que los archivos RR y Actigraph contienen datos de alta dimensionalidad, fue necesario procesarlos de forma separada y luego combinarlos en una única tabla para su posterior análisis.

Los siguientes usuarios presentan anomalías en sus datos:

- En el usuario 1, aparecen registros duplicados debido a una interrupción del sueño, en la que se despertó durante la noche y, después de 30 minutos, regresó a la cama. Este evento provoca que las variables asociadas al tiempo total en cama y al tiempo dormido presenten valores mínimos para este usuario. Para evitar duplicidades en el análisis, se consolidaron ambos registros en uno solo.
- En el usuario 11, no se dispone de datos en la carpeta correspondiente al sueño, lo que afecta al análisis de sus patrones circadianos. Dado que este conjunto de datos es crucial para el estudio, el usuario 11 fue excluido de los análisis que lo combinan con otras fuentes de datos.

- En el usuario 18, se detectó un error en los datos demográficos, donde la edad está registrada como 0. Para corregir este fallo, se decidió sustituir el valor por la media de edad de los demás usuarios.
- En el usuario 21, no aparecen registros en las pruebas de saliva, por lo que no se dispone de sus datos de cortisol ni melatonina. Esto limita su análisis en relación con los aspectos hormonales, por lo que fue excluido de los estudios que combinan este archivo con otros.

En el conjunto de datos RR, la variable que mide el intervalo entre latidos cardíacos (IBI) muestra un valor mínimo constante de 0.250 segundos en muchos usuarios, lo cual podría deberse a limitaciones en algunos dispositivos para detectar intervalos más cortos. Además, se encontraron valores máximos excesivamente altos, posiblemente debidos a errores en la detección de latidos. Debido a estas inconsistencias y a que ya disponemos de información de la frecuencia cardíaca en el archivo Actigraph, se decidió descartar esta variable para el análisis.

Debido a la variabilidad en los horarios de inicio y finalización del experimento entre los distintos usuarios, se ha llevado a cabo una sincronización de los datos, ajustando el periodo de observación a un intervalo común de aproximadamente 24 horas. Esta sincronización asegura una base de comparación uniforme para los parámetros del modelo FMM, permitiendo una evaluación precisa y coherente entre los usuarios y garantizando una representación adecuada del ciclo circadiano.

4.2 Resultados de los modelos paramétricos

Aplicaremos el modelo FMM a la medida wMvFC calculada para los 22 usuarios del conjunto de datos MMASH. Los parámetros obtenidos para cada usuario, así como el coeficiente de determinación R^2 , se detallan en la Tabla 4.

Usuario	M	A	Alpha	Beta	Omega	R^2
user_1	17.434	20.094	2.030	0.099	0.525	0.498
user_10	21.390	19.949	1.981	5.141	0.326	0.482
user_11	14.611	18.605	1.851	5.429	0.257	0.426
user_12	15.436	19.539	2.012	0.371	0.507	0.573
user_13	19.490	22.387	2.712	1.324	0.525	0.682
user_14	12.910	17.088	1.532	6.078	0.419	0.522
user_15	19.252	23.667	2.433	0.759	0.481	0.641
user_17	19.109	22.318	2.610	1.301	0.394	0.684
user_18	13.456	16.203	1.190	5.268	0.193	0.284
user_19	18.535	21.668	2.902	1.665	0.385	0.646
user_2	18.319	14.567	2.787	1.342	0.413	0.429
user_20	18.554	22.852	2.917	1.517	0.241	0.601
user_21	12.854	17.219	2.537	0.823	0.239	0.447
user_22	15.447	20.347	2.133	0.368	0.347	0.528
user_3	16.526	19.939	1.516	6.183	0.519	0.492
user_4	15.008	18.757	2.459	0.970	0.354	0.582
user_5	11.902	15.133	2.829	1.322	0.226	0.306
user_6	16.916	22.216	1.431	5.366	0.365	0.567
user_8	15.701	13.333	0.617	4.727	0.381	0.419
user_9	9.130	11.820	0.999	5.118	0.426	0.426
user_16	21.255	19.714	6.171	4.185	0.751	0.598
user_7	22.226	24.034	5.320	3.824	0.867	0.746

Tabla 4: Parámetros del modelo FMM para los usuarios.

En la Tabla 4 se observa una notable variabilidad entre los distintos usuarios. A continuación, se aplicarán varios métodos de agrupamiento para clasificar a los usuarios en diferentes grupos. Estos usuarios serán clasificados según los parámetros obtenidos del modelo FMM. Los resultados de cada grupo serán analizados en detalle.

Dado que los parámetros α y β son circulares, la distancia euclídea no es adecuada. Por ello, se utilizaron las transformaciones seno y coseno para representarlas correctamente en un espacio cartesiano, respetando su naturaleza cíclica. Los métodos de agrupamiento aplicados incluyen el algoritmo K-medias, el agrupamiento jerárquico utilizando cuatro métodos (*single*, *complete*, *average* y *ward*, ver [26]), así como modelos de mezcla gaussiana (GMM). Se consideraron entre 2 y 10 posibles grupos, y se compararon los valores de silueta [27] obtenidos para cada configuración, como se

detalla en la Tabla 5.

Método	2	3	4	5	6	7	8	9	10
K-medias	0.304	0.289	0.288	0.366	0.331	0.327	0.295	0.252	0.216
GMM	0.180	0.244	0.188	0.303	0.332	0.287	0.285	0.251	0.207
ward	0.304	0.306	0.303	0.325	0.296	0.300	0.260	0.217	0.202
single	0.469	0.349	0.251	0.186	0.019	0.161	0.109	0.128	0.099
complete	0.469	0.237	0.241	0.343	0.329	0.279	0.236	0.201	0.191
average	0.469	0.299	0.259	0.315	0.287	0.242	0.258	0.225	0.208

Tabla 5: Métricas de silueta para diferentes métodos de clustering con los parámetros del modelo FMM.

Se observa que la métrica de silueta alcanza su valor máximo con 2 grupos cuando se utiliza el método de clustering jerárquico (ver [26]), independientemente del criterio de enlace empleado (single, complete o average). Aunque el mejor rendimiento en términos de la métrica de silueta se obtiene con 2 grupos, el análisis con el modelo de K-means utilizando 5 clusters puede ser valioso para identificar y distinguir un mayor número de patrones o casos específicos en los datos, ofreciendo una perspectiva adicional sobre la estructura subyacente.

A continuación, analizaremos los métodos de clustering que presentan los valores resaltados en negrita en la Tabla 5.

4.2.1 Clustering jerárquico con 2 clusters

Con el objetivo de visualizar los diferentes grupos, proyectamos los datos de la Tabla 4 en dos componentes principales que capturan cerca del 75% de la información. Cada color representa uno de los dos grupos obtenidos con los métodos jerárquicos y los parámetros de la Tabla 4.

Esta visualización se muestra en la Figura 6.

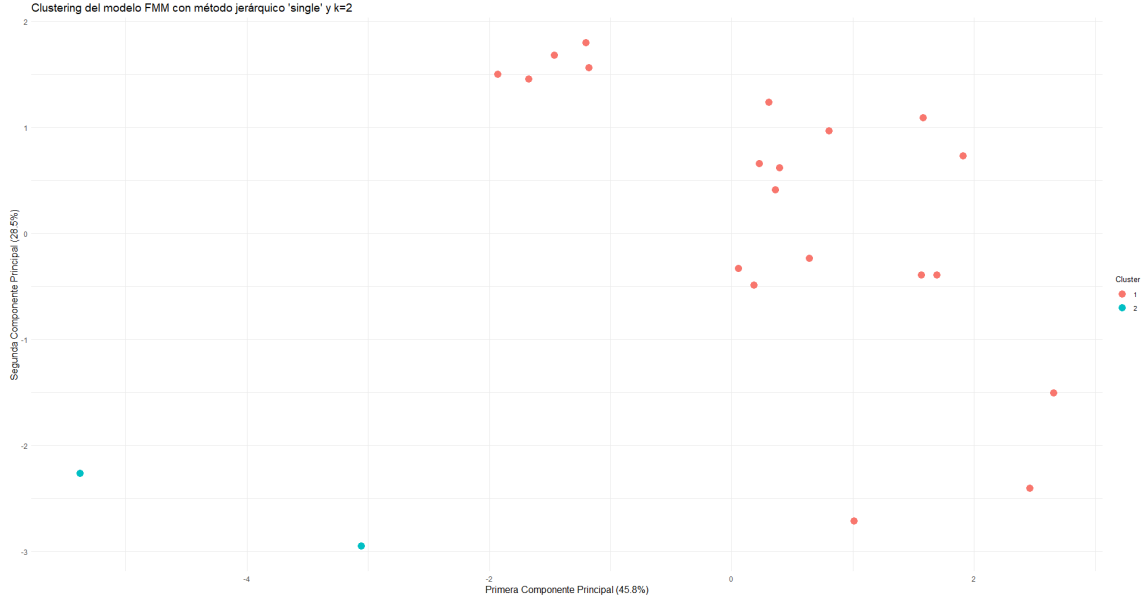


Figura 6: Proyección de los clusters obtenidos con el método jerárquico en dos componentes principales. Cada grupo es representado con un color diferente.

Únicamente dos usuarios, el 7 y el 16, son clasificados en el grupo 2. A continuación, en la Tabla 6 se muestran los parámetros de la Tabla 4 para estos dos usuarios diferenciados.

Usuario	Cluster	M	A	Alpha	Beta	Omega	R^2
user_16	2	21.255	19.714	6.271	4.185	0.751	0.598
user_7	2	22.226	24.034	5.320	3.824	0.867	0.746

Tabla 6: Parámetros del modelo FMM para los usuarios del cluster 2 obtenido con método jerárquico *single*.

Estos dos usuarios se caracterizan por tener valores altos de ω , lo que sugiere que las ondas ajustadas son predominantemente sinusoidales. También, sus valores de α son superiores a π y están cerca de 2π . Los valores de β también son superiores a π . Además, los valores de M y A son de los más altos obtenidos en el análisis. Este cluster se destaca por tener, en promedio, el mejor ajuste a los datos. Estas características indican que las ondas en este cluster alcanzan valores altos y su forma será sinusoidal, lo cual contribuye a un buen ajuste del modelo.

En cuanto al otro grupo, los datos muestran una alta variabilidad en los parámetros.

Para identificar diferentes patrones, hemos seleccionado un método con un mayor número de clusters, aunque con un menor valor de silueta. El método elegido es el de K-medias con 5 grupos, que proporciona el segundo valor más alto de silueta obtenido.

Es importante señalar que un número elevado de clusters en un conjunto de datos con dimensiones reducidas puede afectar negativamente el valor de la silueta. A pesar de que un mayor número de clusters puede revelar patrones adicionales, el ajuste se ve penalizado debido al bajo número de usuarios (22 en total).

4.2.2 Método K-medias con 5 clusters

Con el objetivo de visualizar los diferentes grupos, proyectamos los datos de la Tabla 4 en dos componentes principales que capturan cerca del 75 % de la información. Cada color representa uno de los cinco grupos obtenidos con el método de K-medias. Esta visualización se muestra en la Figura 7.

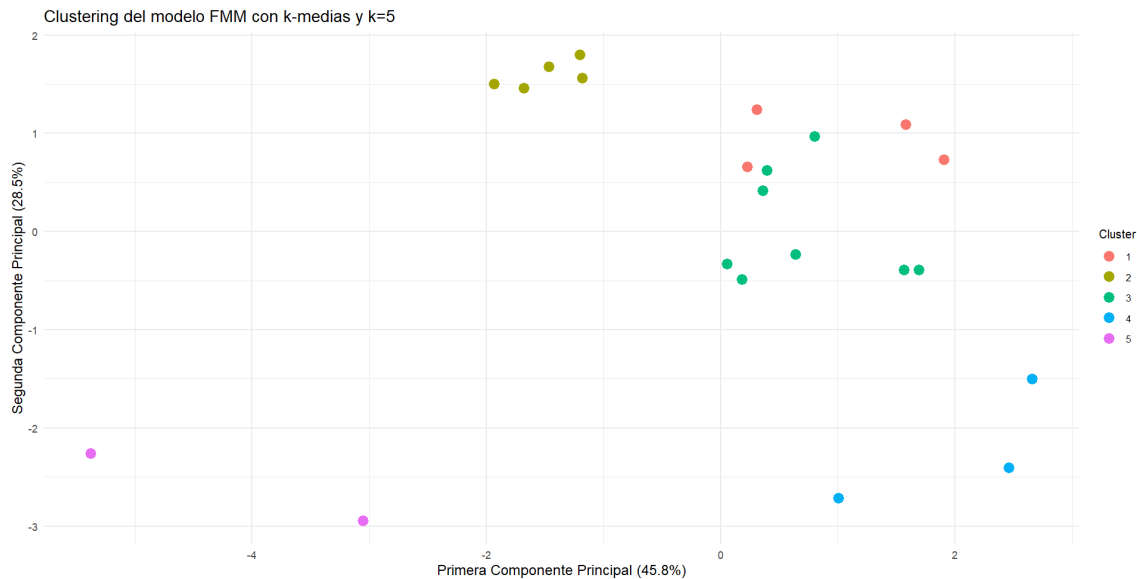


Figura 7: Proyección de los clusters obtenidos con el método K-medias ($k=5$) en dos componentes principales. Cada grupo es representado con un color diferente.

Se observa que los grupos están bien diferenciados en estas dos dimensiones, a excepción de los grupos 1 y 3, que muestran cierta superposición en algunos puntos.

La Tabla 7 muestra los parámetros del modelo FMM de la Tabla 4 añadiendo un indicador del grupo al que pertenecen.

Usuario	Cluster	M	A	Alpha	Beta	Omega	R^2
user_2	1	18.319	14.567	2.787	1.342	0.413	0.429
user_21	1	12.854	17.219	2.537	0.823	0.239	0.447
user_4	1	15.008	18.757	2.459	0.970	0.354	0.582
user_5	1	11.902	15.133	2.829	1.322	0.226	0.306
user_13	2	19.490	22.387	2.713	1.324	0.525	0.682
user_15	2	19.252	23.667	2.433	0.759	0.481	0.641
user_17	2	19.109	22.318	2.610	1.301	0.394	0.684
user_19	2	18.535	21.668	2.902	1.665	0.385	0.646
user_20	2	18.554	22.852	2.917	1.517	0.241	0.601
user_1	3	17.434	20.094	2.030	0.099	0.525	0.498
user_10	3	21.390	19.949	1.981	5.141	0.326	0.482
user_11	3	14.611	18.605	1.851	5.429	0.257	0.426
user_12	3	15.436	19.539	2.012	0.371	0.507	0.573
user_14	3	12.910	17.088	1.532	6.078	0.419	0.522
user_22	3	15.447	20.347	2.133	0.368	0.347	0.528
user_3	3	16.526	19.939	1.516	6.183	0.519	0.492
user_6	3	16.916	22.216	1.431	5.366	0.365	0.567
user_18	4	13.456	16.203	1.190	5.268	0.193	0.284
user_8	4	15.701	13.333	0.617	4.727	0.381	0.419
user_9	4	9.130	11.820	0.999	5.118	0.426	0.426
user_16	5	21.255	19.714	6.271	4.185	0.751	0.598
user_7	5	22.226	24.034	5.320	3.824	0.867	0.746

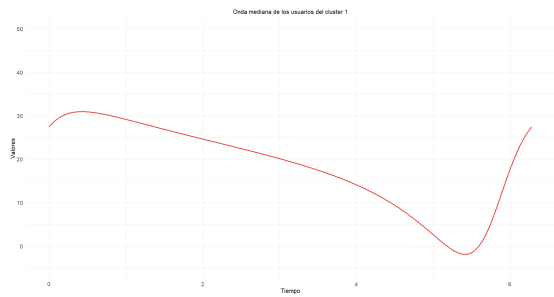
Tabla 7: Parámetros del modelo FMM para los usuarios y clusters de K-medias.

Los clusters identificados muestran las siguientes tendencias:

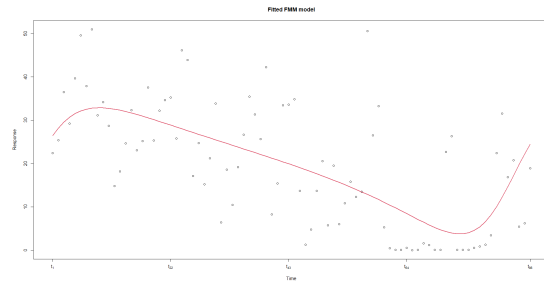
- **Cluster 1:** Este grupo presenta valores moderados de M y A . Los valores de α , cercanos a 2.5, junto con valores cercanos a 1 para β y bajos para ω , generan una onda asimétrica y apuntada. La onda alcanza su máximo al inicio del ciclo y desciende de manera controlada hasta las horas nocturnas, donde experimenta un crecimiento abrupto. La mayoría de los usuarios en este cluster tienen un coeficiente R^2 moderado-bajo, lo que sugiere que el modelo se ajusta de manera razonable a sus datos.

- **Cluster 2:** Los usuarios de este cluster presentan una notable consistencia en sus parámetros. Los valores de M y A son superiores en comparación con el cluster 1, lo que indica amplitudes más altas. Los parámetros α , β y ω son similares a los del cluster anterior, por lo que las formas de sus ondas serán similares, aunque debido a los valores más altos de M y A , la onda alcanzará valores máximos superiores. Además, este grupo tiene un R^2 consistentemente elevado, lo que indica un buen ajuste del modelo a los datos de los usuarios.
- **Cluster 3:** Este es el cluster con mayor número de usuarios y se caracteriza por valores moderados de M , A . Los valores de α se encuentran entre 1.5 y 2. Los valores cercanos a 0 o muy elevados cercanos a 2π de β y entre 0.3 y 0.5 de ω . Estos parámetros describen una onda asimétrica pero suavizada, que disminuye progresivamente en las horas de menor actividad. Sin embargo, el aumento posterior ocurre de forma abrupta. Los usuarios de este grupo tienen un coeficiente R^2 moderado, superior que en el cluster 1.
- **Cluster 4:** Los usuarios en este cluster presentan los valores positivos más bajos de α , lo que implica una fase temprana de la onda. Los valores de β son los más altos, sugieren que este grupo presenta ondas simétricas o casi simétricas. Los otros parámetros (M , A , y ω) muestran gran variación entre los usuarios. Además, la variabilidad en ω indica que este cluster incluye tanto ondas sinusoidales como no sinusoidales. El R^2 es relativamente bajo, sugiriendo que el modelo tiene dificultades para capturar completamente los patrones de este grupo.
- **Cluster 5:** Este cluster agrupa a los dos usuarios que fueron separados por el clustering jerárquico, tal y como se explicó en la sección anterior.

En las Figuras 8-12 se presentan dos imágenes para cada cluster. A la izquierda, se muestra la gráfica generada a partir de la mediana de los parámetros de cada grupo, con el eje X que varía de 0 a 2π . A la derecha, se presenta el modelo FMM ajustado a un usuario específico dentro de dicho grupo, donde el eje X representa el número de valores utilizados para la representación.

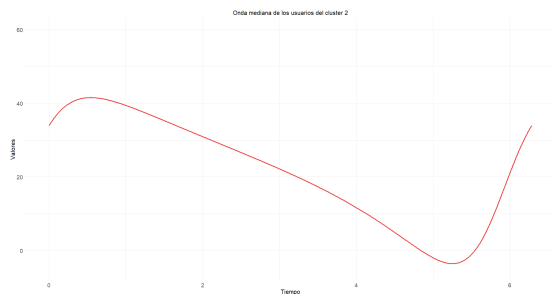


(a) Onda FMM de la mediana.

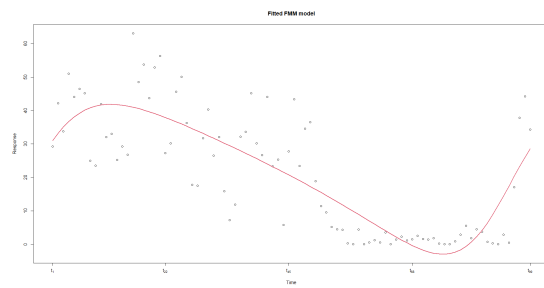


(b) Usuario 2 del cluster 1. $R^2 = 0,429$.

Figura 8: (a) Onda FMM obtenida con la mediana de los parámetros del cluster 1. (b) Modelo FMM ajustado para el usuario 2 de dicho cluster.

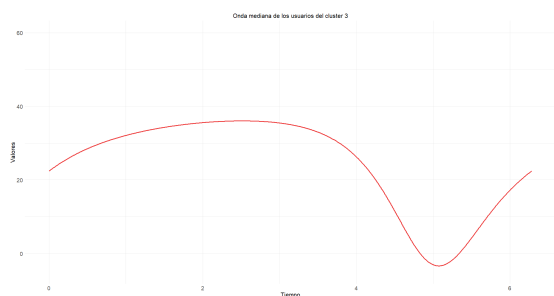


(a) Onda FMM de la mediana.

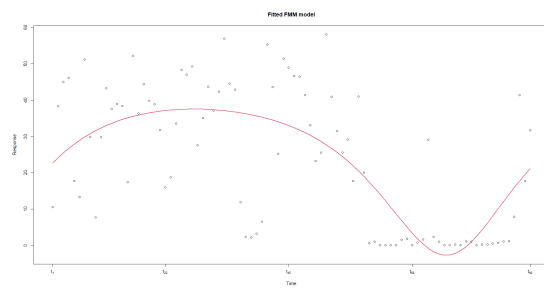


(b) Usuario 13 del cluster 2. $R^2 = 0,682$.

Figura 9: (a) Onda FMM obtenida con la mediana de los parámetros del cluster 2. (b) Modelo FMM ajustado para el usuario 13 de dicho cluster.

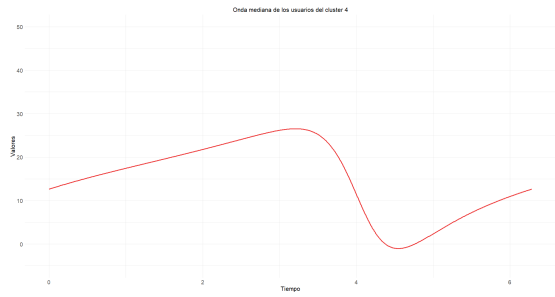


(a) Onda FMM de la mediana.

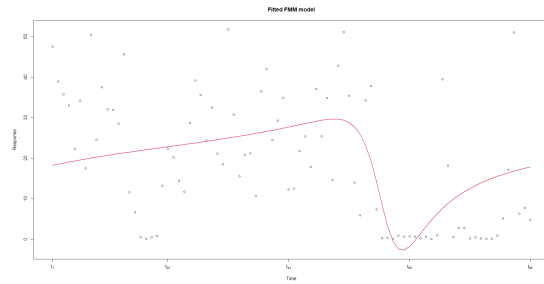


(b) Usuario 1 del cluster 3. $R^2 = 0,498$.

Figura 10: (a) Onda FMM obtenida con la mediana de los parámetros del cluster 3. (b) Modelo FMM ajustado para el usuario 1 de dicho cluster.

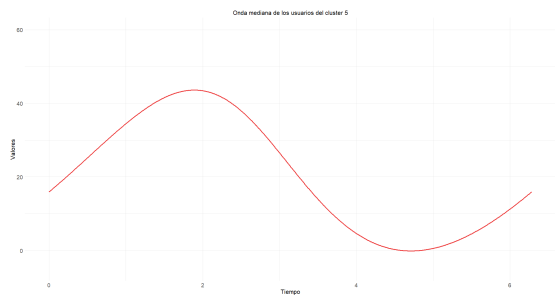


(a) Onda FMM de la mediana.

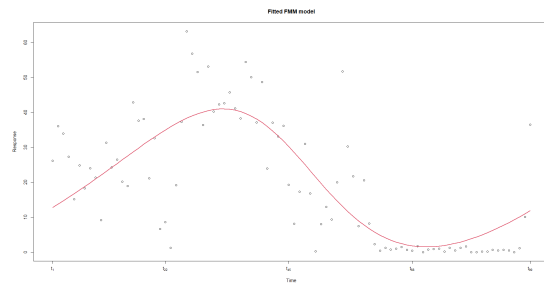


(b) Usuario 18 del cluster 4. $R^2 = 0,284$.

Figura 11: (a) Onda FMM obtenida con la mediana de los parámetros del cluster 4. (b) Modelo FMM ajustado para el usuario 18 de dicho cluster.



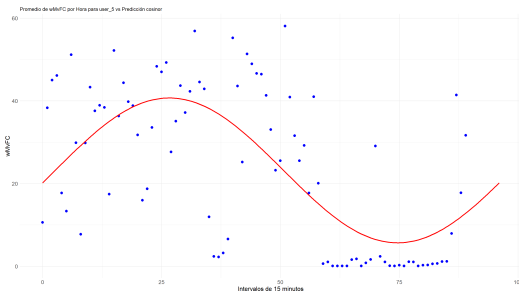
(a) Onda FMM de la mediana.



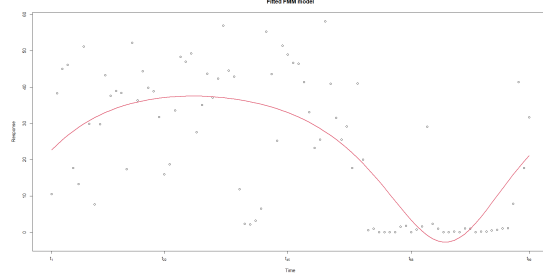
(b) Usuario 16 del cluster 5. $R^2 = 0,598$.

Figura 12: (a) Onda FMM obtenida con la mediana de los parámetros del cluster 5. (b) Modelo FMM ajustado para el usuario 16 de dicho cluster.

A continuación, se presenta una comparación con el modelo cosinor aplicado al usuario 1 (Figura 13) y al usuario 8 (Figura 14).



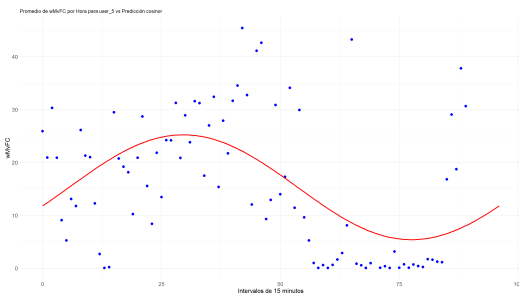
(a) Modelo cosinor.



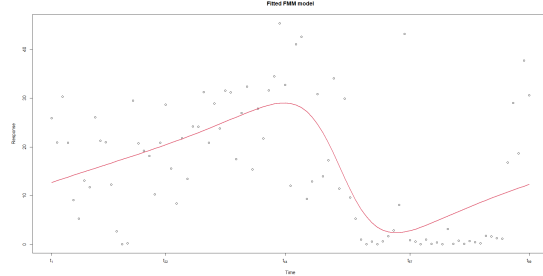
(b) Modelo FMM.

Figura 13: Comparativa de los modelos cosinor y FMM para el usuario 1.

Se observa que la onda del modelo cosinor comienza a descender prematuramente, sin tener en cuenta las observaciones con mayor amplitud al final de la tarde y sin aproximarse a los valores mínimos de la onda. En contraste, la onda del modelo FMM muestra una ligera concavidad durante los períodos de mayor actividad y desciende posteriormente hasta alcanzar los valores mínimos. Sin embargo, el modelo FMM presenta una limitación, ya que los valores predichos incluyen valores negativos.



(a) Modelo cosinor.



(b) Modelo FMM.

Figura 14: Comparativa de los modelos cosinor y FMM para el usuario 8.

Se observa que la onda del modelo cosinor comienza a descender prematuramente, sin tener en cuenta las observaciones con mayor amplitud al final de la tarde y sin alcanzar los valores mínimos. En contraste, el modelo FMM muestra una tendencia ascendente hacia estos valores de mayor amplitud antes de experimentar un descenso más pronunciado, lo que permite alcanzar los valores mínimos.

4.3 Resultados obtenidos de los métodos no paramétricos

En primer lugar, abordaremos el análisis de componentes principales funcionales (ACPF), y posteriormente los índices no paramétricos.

4.3.1 Resultados del Análisis de Componentes Principales Funcionales (ACPF)

Siguiendo la metodología del apartado anterior, se calculó la medida wMvFC para todos los usuarios. Utilizando la librería “fdapace” del software R, se realizó un análisis de componentes principales funcionales considerando 4 componentes.

Estas cuatro componentes explican el 48 %, 40 %, 9 % y 2 % de la varianza, respectivamente. En las Figuras 15,16 se representa la función media $\rho(t)$ junto con las 4 autofunciones obtenidas.

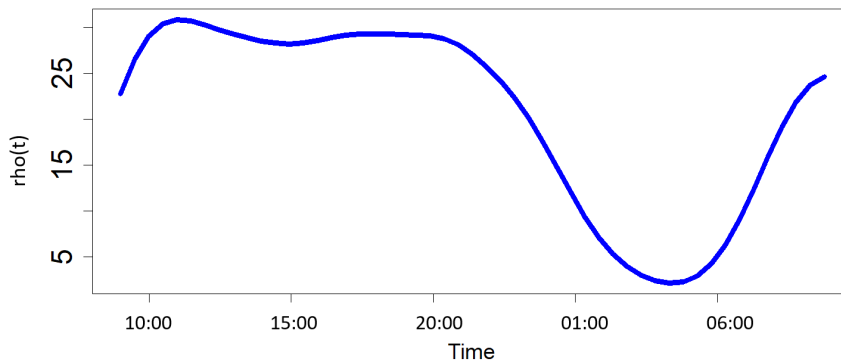


Figura 15: Función media $\rho(t)$.

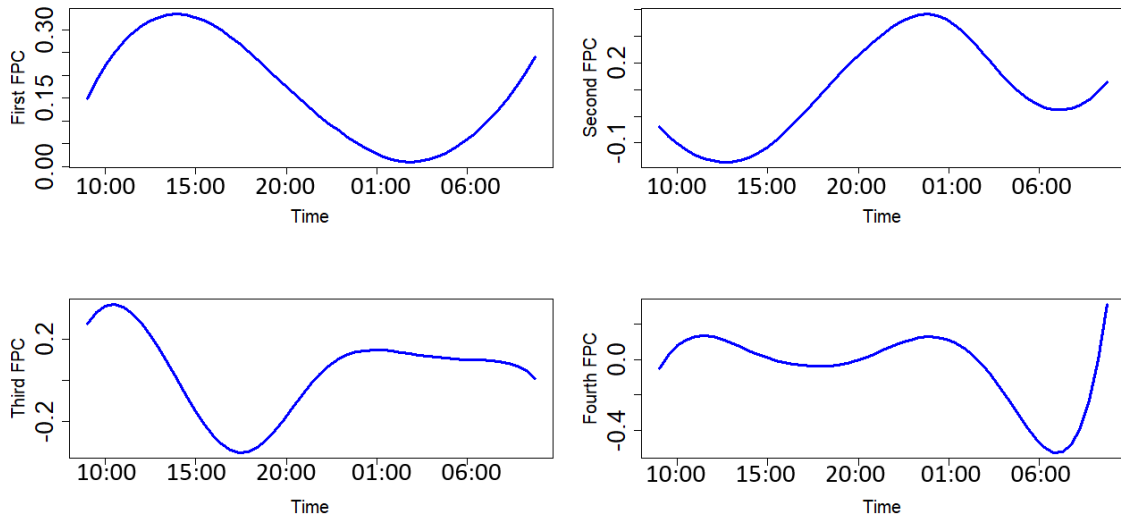


Figura 16: Cuatro primeras autofunciones.

La primera autofunción es positiva a lo largo de todo el periodo analizado, lo que nos permite interpretarla como la evolución de la media temporal.

La segunda autofunción presenta valores negativos, tal como cabría esperar, dado que debe ser ortonormal respecto a la primera. Esta componente puede asociarse a la evolución de la varianza a lo largo del tiempo, mostrando valores más elevados al inicio de la fase del sueño, donde se observa mayor variabilidad en función de cuándo los usuarios inician esta fase.

La tercera autofunción refleja la variabilidad entre los usuarios durante las primeras horas del día, que corresponden a los periodos de mayor actividad, y se sitúa alrededor de 0 durante las últimas horas de la tarde y la noche.

Finalmente, la cuarta autofunción se mantiene cerca de 0 hasta las últimas horas de la noche, momento en que desciende, toma valores negativos y, al llegar el despertar, aumenta drásticamente, marcando el fin del experimento y capturando la variabilidad observada entre los usuarios durante esta fase.

La Figura 17 muestra la aproximación de los datos iniciales utilizando la Fórmula 1.

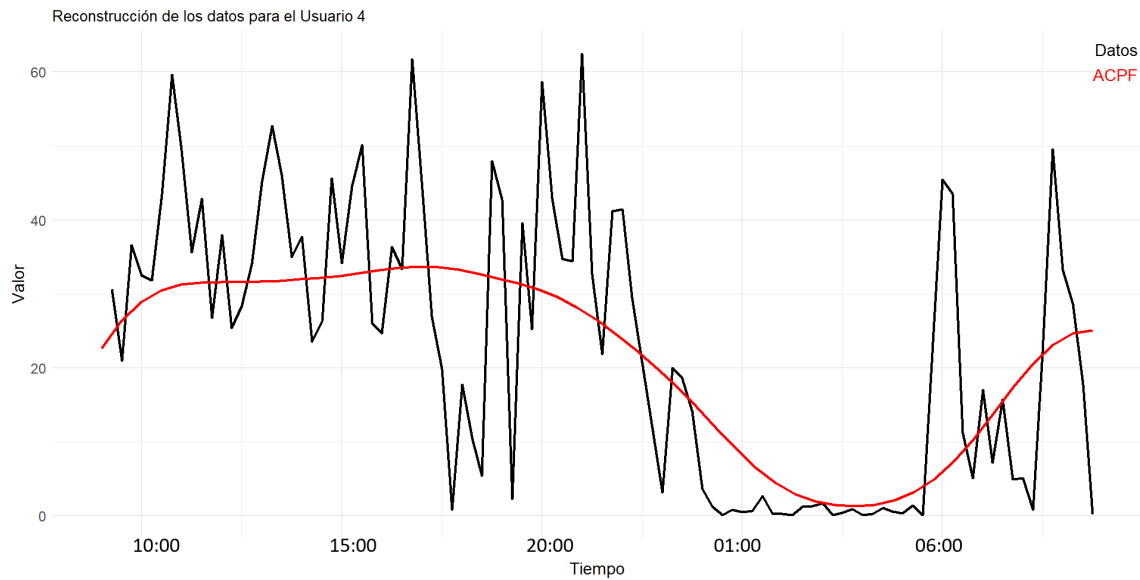


Figura 17: Aproximación ACPF para el usuario 4.

La aproximación recoge la tendencia principal de los datos.

Para concluir este apartado, se llevará a cabo un análisis de agrupamiento (clustering) de las series temporales utilizando la aproximación de los datos de todos los usuarios. Para este propósito, se empleará el algoritmo de deformación dinámica del tiempo (DTW, por sus siglas en inglés, *Dynamic Time Warping*), el cual es una técnica ampliamente utilizada para medir la similitud entre dos series temporales, incluso en situaciones donde las series presentan variaciones en velocidad o desfases temporales. El principio central del DTW es encontrar una alineación óptima entre las series, permitiendo distorsiones temporales en el proceso. Esto facilita que un único punto en una serie temporal pueda alinearse con múltiples puntos de la otra serie, lo que permite una comparación más precisa entre ambas.

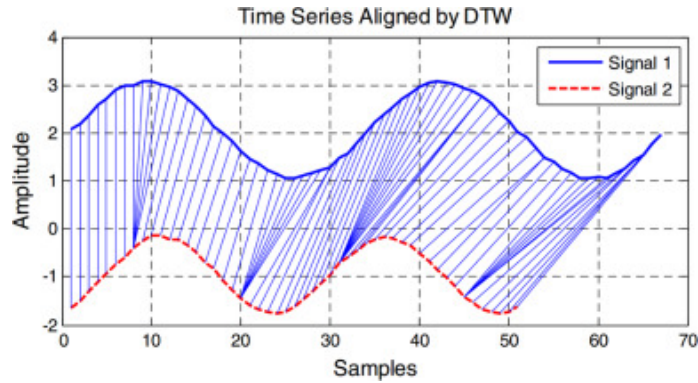


Figura 18: Ejemplo de DTW para dos series temporales.

Siguiendo la metodología establecida en el apartado anterior, se probarán diferentes métodos de agrupamiento sobre la matriz de distancias obtenida mediante DTW.

La Tabla 8 muestra los valores del coeficiente de silueta para los distintos métodos de clustering evaluados.

Método	2	3	4	5	6	7	8	9	10
K-medias	0.489	0.249	0.218	0.198	0.133	0.128	0.127	0.115	0.088
GMM	0.489	0.249	0.218	0.197	0.207	0.197	0.163	0.166	0.156
ward	0.313	0.209	0.224	0.205	0.151	0.142	0.173	0.151	0.175
single	0.463	0.432	0.261	0.122	0.049	0.030	0.131	0.079	0.088
complete	0.313	0.280	0.217	0.205	0.202	0.187	0.167	0.177	0.205
average	0.452	0.401	0.261	0.226	0.207	0.202	0.149	0.142	0.158

Tabla 8: Métricas de silueta para diferentes métodos de clustering con la distancia DTW y ACPF.

Se observa que los mejores resultados se obtienen al agrupar a los usuarios en dos clusters, utilizando los métodos GMM y K-medias. Aunque para un mayor número de grupos el siguiente mejor resultado corresponde al método jerárquico *single* con 3 clusters, este método produce una segmentación inadecuada al asignar un único usuario a un cluster, otro al segundo y a todos los demás al tercero. Debido a esta distribución poco representativa, no se considerará en el análisis posterior.

La Tabla 9 presenta la distribución de los usuarios en cada uno de los grupos identificados.

Cluster 1	Cluster 2
user_10	user_1
user_11	user_12
user_18	user_13
user_21	user_14
user_5	user_15
user_6	user_16
user_8	user_17
user_9	user_19
	user_2
	user_20
	user_22
	user_3
	user_4
	user_7

Tabla 9: Usuarios en clusteres 1 y 2 usando K-medias, DTW y ACPF.

La Figura 19 representa la aproximación obtenida mediante la Fórmula 1 para todos los usuarios y los clusteres representados por diferentes colores.

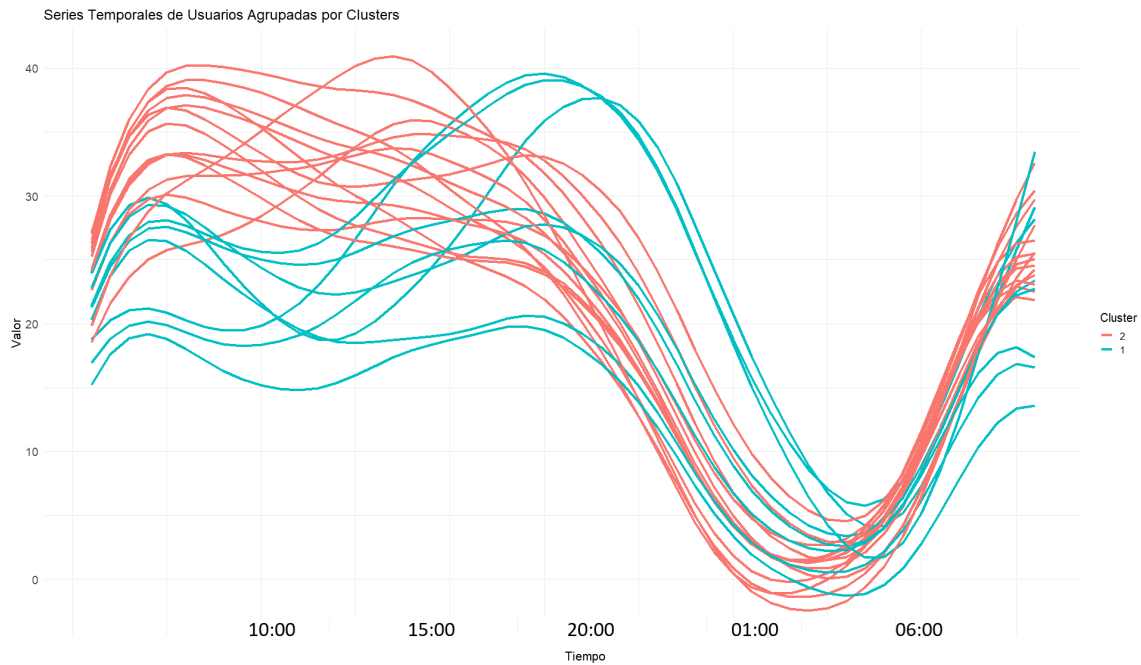
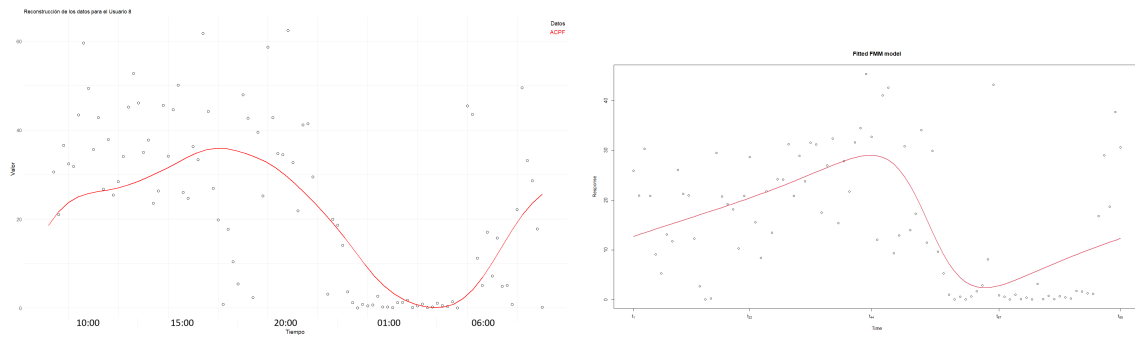


Figura 19: Aproximación para todos los usuarios clasificados en dos grupos con K-medias, DTW y ACPF.

En la Figura 19 se observan claramente los dos grupos identificados. El grupo 1 (en rojo) incluye a aquellos usuarios cuya aproximación muestra valores mínimos más bajos en comparación con el grupo 2 (en azul). Adicionalmente, los usuarios del grupo 1 comienzan con valores más altos que disminuyen desde los primeros instantes del tiempo, mientras que los del grupo 2 tienden a mantener sus valores estables o incluso aumentan, alcanzando las horas de descanso a un tiempo más tardío.

Finalmente, en las Figuras 20 y 21 representaremos la aproximación obtenida con ACPF y el ajuste del modelo FMM para un usuario de cada cluster, el usuario 8 para el cluster 1 y el usuario 1 para el cluster 2.

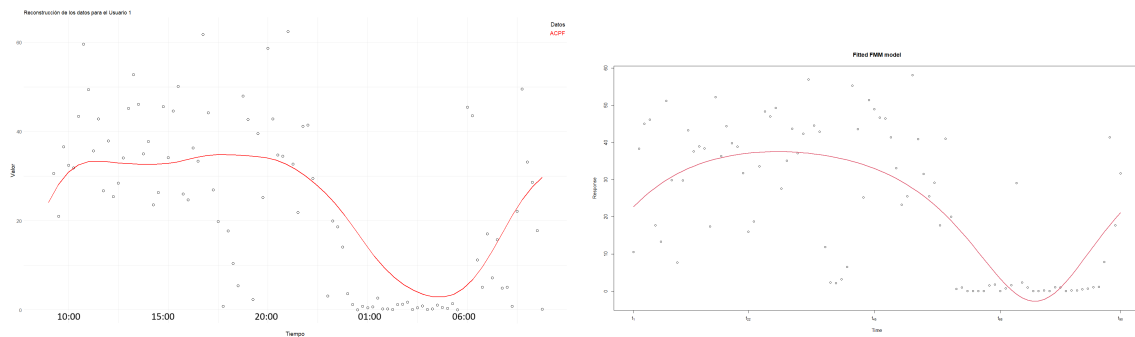


(a) Aproximación ACPF del usuario 8.

(b) Ajuste FMM para el usuario 8.

Figura 20: Aproximación ACPF y ajuste FMM para el usuario 8 del cluster 1.

Se aprecia que la aproximación obtenida mediante ACPF para el usuario 8 es más precisa, ya que refleja de manera adecuada la estabilización durante los periodos nocturnos, además de capturar la tendencia ascendente de los datos hasta las horas de menor actividad.



(a) Aproximación ACPF del usuario 1.

(b) Ajuste FMM para el usuario 1.

Figura 21: Aproximación ACPF y ajuste FMM para el usuario 1 del cluster 2.

Por otro lado, para el usuario 1 del cluster 2, el modelo FMM ofrece un mejor ajuste. En este caso, la aproximación obtenida mediante ACPF se mantiene constante durante los periodos de vigilia y no refleja adecuadamente la tendencia cóncava observada en los datos. Asimismo, no alcanza los valores mínimos reales del usuario, mientras que el modelo FMM, aunque produce valores negativos, sí logra capturar estos mínimos.

4.3.2 Resultados de los índices no paramétricos y cuestionarios

El objetivo de este apartado es calcular los índices no paramétricos y analizar cómo se relacionan los resultados obtenidos con otras variables disponibles en el conjunto de datos.

Iniciaremos el análisis calculando los índices no paramétricos de la frecuencia cardíaca, para lo que utilizaremos un enfoque de ventana deslizante. Este método optimiza el proceso de cálculo de acumulados al evitar la repetición de sumas innecesarias. En lugar de calcular la suma completa desde cero en cada paso, se realiza una suma inicial para un intervalo, por ejemplo, de 5 horas, y para el siguiente acumulado simplemente se añade el valor siguiente y se resta el primer valor de la suma anterior. De esta manera, se reduce significativamente el tiempo de cálculo al evitar la repetición de sumas desde el inicio en cada paso.

Este método resulta particularmente eficiente cuando se trabaja con grandes volúmenes de datos en intervalos cortos de tiempo, como los valores minuto a minuto de las variables consideradas. Al aplicar la ventana deslizante, se mejora la eficiencia del cálculo y se permite obtener resultados precisos de manera más rápida, sin necesidad de recalcular completamente cada acumulado para cada nueva ventana de tiempo. En nuestro caso, se logró reducir el número de operaciones de 700,000 sumas a 1,500. Una vez obtenidos los acumulados, se identifican los valores máximos y mínimos. Para las variables temporales, se calcula el mínimo o máximo correspondiente y, posteriormente, se centran los valores para obtener el valor medio.

Los índices de la frecuencia cardíaca se presentan en la Tabla 10. Los valores de amplitud se expresan en latidos por minuto, mientras que los de tiempo indican la hora de la medida, comenzando entre las 9:00 y las 10:00. Si los valores de tiempo superan 24, se refieren a horas del día siguiente, por ejemplo, 25 corresponde a la 1:00 de la madrugada. Los valores de amplitud absoluta (AA) y amplitud relativa (RA) se encuentran en la Tabla 11.

Hemos identificado una relación entre la amplitud relativa, la amplitud absoluta y la variable de melatonina tras un ajuste que explicamos a continuación. Como se observó previamente en el análisis de la melatonina (ver Figura 5), la distribución de los datos no presentó variaciones significativas entre las mediciones realizadas antes de dormir y al despertar. Para solventar esta posible falta de variación, se calculó una proporción de cambio mediante la siguiente fórmula:

$$\text{melatonina_ajustada} = \frac{\text{melatonina_AD} - \text{melatonina_DD}}{\text{máx}(\text{melatonina}) - \text{mín}(\text{melatonina})}$$

donde AD y DD significa antes de dormir y despues de dormir, respectivamente.

Usuario	VM5	TM5	VL5	TL5	VM10	TM10	VL10	TL10
user_1	96.14	16.17	67.49	28.98	93.62	18.48	74.78	26.93
user_2	79.72	23.12	68.23	14.37	75.83	21.47	70.62	16.87
user_3	79.89	13.23	58.62	29.73	74.12	16.08	65.58	28.42
user_4	81.28	15.60	56.79	26.88	79.11	14.90	61.83	26.25
user_5	90.53	15.45	59.69	28.25	86.50	14.58	62.96	27.53
user_6	83.89	11.68	58.53	26.80	78.14	14.18	62.78	25.67
user_7	112.58	17.60	62.41	28.68	100.67	14.50	70.08	27.05
user_8	82.36	18.37	50.34	29.58	79.56	15.47	56.79	26.97
user_9	103.15	18.67	59.00	27.55	98.96	16.15	68.27	26.17
user_10	75.87	22.12	62.26	27.88	73.71	20.02	67.80	26.77
user_11	93.99	12.98	56.42	29.78	86.95	14.88	57.42	27.50
user_12	86.18	17.28	55.72	26.42	82.43	15.47	65.42	25.67
user_13	83.68	12.70	53.00	28.07	80.38	14.45	58.88	26.18
user_14	84.36	12.18	61.53	28.77	81.91	14.68	66.96	26.50
user_15	74.64	20.75	51.25	28.62	72.17	19.97	58.72	26.83
user_16	91.07	18.65	56.44	26.60	85.68	16.07	66.09	26.42
user_17	79.79	15.85	58.40	27.25	78.64	14.13	63.98	24.75
user_18	82.46	11.92	65.69	29.10	79.64	14.42	68.87	26.97
user_19	97.73	13.12	68.89	28.80	96.68	15.62	73.56	28.12
user_20	90.96	17.18	61.05	29.63	88.10	15.33	62.81	27.25
user_21	83.75	20.52	60.82	26.73	78.80	18.82	66.57	27.05
user_22	101.14	21.52	56.71	30.23	83.97	18.73	62.89	27.93

Tabla 10: Índices no paramétricos para la Frecuencia Cardiaca (1).

Usuario	RA	AA
user_1	0.1622	26.1363
user_2	0.1934	26.7123
user_3	0.2057	29.2349
user_4	0.1477	20.2394
user_5	0.0959	13.9410
user_6	0.1678	27.7884
user_7	0.1814	27.0503
user_8	0.1287	17.9746
user_9	0.1938	27.2635
user_10	0.0528	7.6044
user_11	0.1167	15.4964
user_12	0.1642	22.3139
user_13	0.1834	26.8156
user_14	0.1435	19.6131
user_15	0.2346	38.2562
user_16	0.2249	29.2148
user_17	0.2530	39.9690
user_18	0.0842	11.4538
user_19	0.2129	30.5247
user_20	0.2052	27.3737
user_21	0.1420	20.3741
user_22	0.1696	20.9264

Tabla 11: Índices no paramétricos para la Frecuencia Cardiaca (2).

La relación encontrada entre la melatonina ajustada y los valores de RA y AA es una correlación negativa de -0.52 para ambos casos. Esto indica que, en nuestra muestra, a medida que aumenta la diferencia en los niveles de melatonina antes y después de dormir, se observa una disminución tanto en la amplitud relativa como en la amplitud absoluta, lo que indica que mayores niveles de melatonina antes de dormir pueden estar vinculados a una menor variabilidad en la frecuencia cardíaca.

La Figura 22 recoge esta relación.

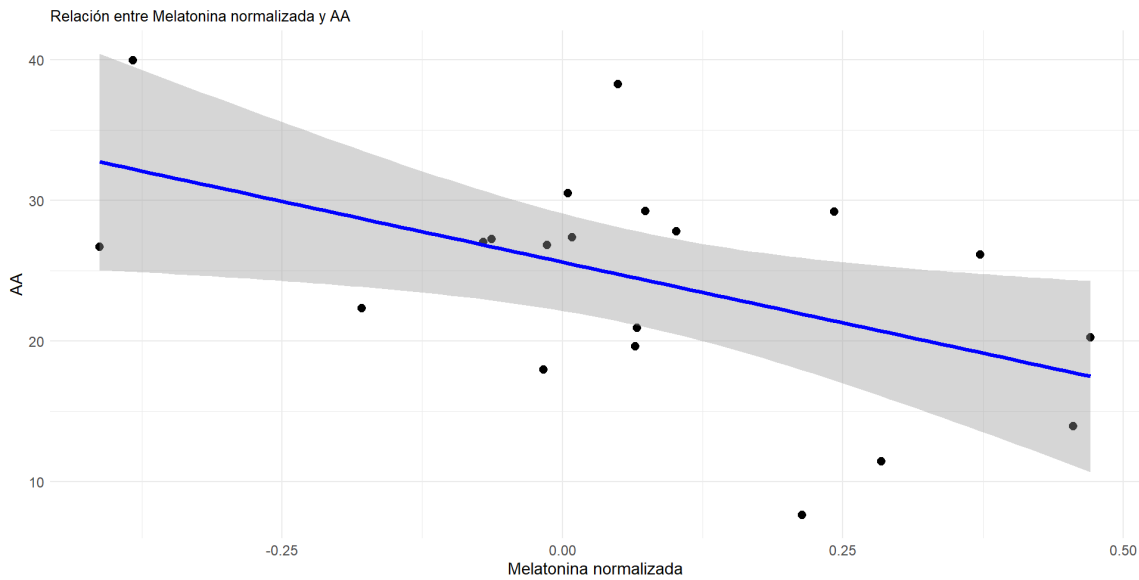


Figura 22: Melatonina ajustada y Amplitud Absoluta.

Se observa una tendencia descendente.

Estudio de los cuestionarios MEQ y PSQI

Hemos intentado relacionar los cuestionarios de MEQ y PSQI (recordamos que recogían información del cronotipo de los usuarios y de como durmieron la noche del experimento respectivamente), con las demás variables sin conseguir ningún resultado relevante.

En contra de lo que podemos pensar, en esta muestra no se ha observado relación entre, el número de despertares en la noche, la duración del sueño, el tiempo en la cama y el índice de Pittsburgh.

Por otro lado, el índice MEQ que indica el cronotipo de los usuarios tampoco se ve relacionado con el tiempo en el que los usuarios alcanzan sus valores máximos o mínimos. Si un usuario es de cronotipo nocturno, debería de tener los valores mínimos de tiempo más altos que los demás.

La ausencia de una relación clara entre los cuestionarios y los datos objetivos puede explicarse por la limitación en la cantidad de días incluidos en el conjunto de datos. Con observaciones de una sola noche, no es posible determinar de manera concluyente

el cronotipo de ningún usuario.

Variabilidad Intradiaria (IV) e Índice de Circadianidad (CI)

A continuación, calculamos el índice de variabilidad intradiaria (IV) para cada usuario, que mide la fragmentación del ritmo circadiano. Los resultados de IV se presentan en la Tabla 12.

Usuario	IV
user_2	0.38666915
user_22	0.35429744
user_18	0.32804389
user_9	0.28026861
user_11	0.26719713
user_10	0.24882554
user_4	0.25164255
user_5	0.25117278
user_8	0.19709934
user_21	0.20449767
user_7	0.17910234
user_6	0.16680715
user_16	0.16580315
user_14	0.14658086
user_13	0.13041827
user_1	0.13921357
user_17	0.12063531
user_12	0.10493990
user_20	0.10816392
user_3	0.09389067
user_15	0.09201119
user_19	0.07824704

Tabla 12: Valores de IV para cada usuario ordenados de mayor a menor.

Observamos que la variabilidad intradiaria varía considerablemente entre los usuarios, desde valores elevados alrededor de 0.38 hasta valores cercanos a 0.08. Las Figuras 23 y 24 representan la frecuencia cardiaca de los usuarios con los valores extremos de IV para observar estas diferencias.

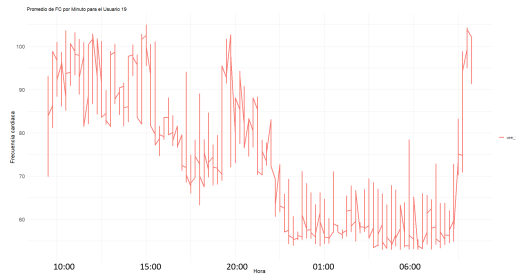


Figura 23: Frecuencia cardiaca por hora para el user_19 con $IV=0.08$.

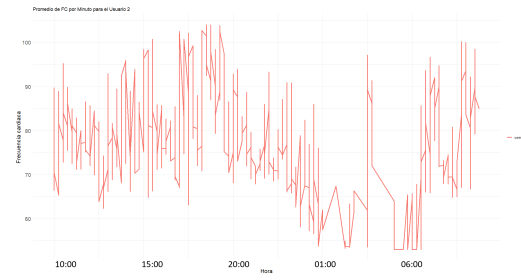


Figura 24: Frecuencia cardiaca por hora para el user_2 con $IV=0.39$.

El usuario 19, con el valor más bajo de IV , muestra una mayor estabilidad en su ritmo circadiano a lo largo del día. Durante los periodos de mayor actividad, su frecuencia cardiaca se mantiene elevada y estable. Se observa una disminución controlada de la frecuencia cardiaca en el periodo posterior a la comida, un patrón esperable debido a la menor actividad física en ese momento, seguido de un nuevo aumento que se estabiliza. En los periodos de menor actividad, la frecuencia cardiaca permanece muy estable y regulada, lo que sugiere un ritmo circadiano coherente y bien definido.

En contraste, el usuario 2, con el valor más alto de IV , presenta una notable inestabilidad incluso durante la noche, donde se registran valores de frecuencia cardiaca tan elevados como en los periodos de mayor actividad. A lo largo del día, este usuario muestra una gran variabilidad en su frecuencia cardiaca, con rápidas fluctuaciones entre aumentos y descensos, sin una pauta de estabilidad definida, lo que refleja un ritmo circadiano fragmentado y desorganizado.

Finalmente, calculamos el índice de circadianidad aplicando la transformada rápida de Fourier a la frecuencia cardiaca de los usuarios. Para este cálculo, agrupamos los datos en intervalos de 60 segundos, tomando el promedio en cada intervalo, y no realizamos una escala adicional. Los resultados obtenidos se muestran en la Tabla 13.

Usuario	CI
user_13	0.81025030
user_7	0.80290703
user_17	0.73803264
user_19	0.71508926
user_6	0.69613263
user_15	0.66262676
user_20	0.63425370
user_12	0.56146847
user_16	0.51503713
user_4	0.52708397
user_3	0.44119314
user_1	0.39383145
user_21	0.39376206
user_22	0.27420346
user_2	0.23677771
user_8	0.34526927
user_9	0.36241147
user_5	0.18918455
user_11	0.18254062
user_18	0.11960733
user_10	0.03605892

Tabla 13: Valores de CI para cada usuario ordenados de mayor a menor.

Observamos como hay una gran variabilidad entre los distintos usuarios, desde aquellos cercanos a 1 con valores alrededor de 0.8, hasta valores del orden de la centésima, casi 0. Las Figuras 25 y 26 representan la frecuencia cardiaca de los usuarios con los valores extremos de CI para observar estas diferencias.

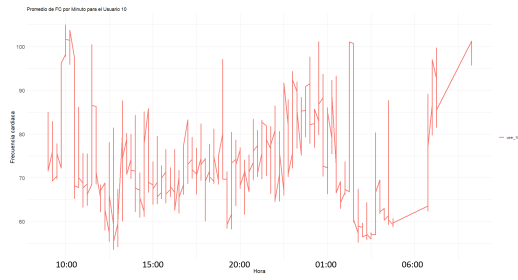


Figura 25: Frecuencia cardiaca por hora para el user_10 con $CI=0.03$.

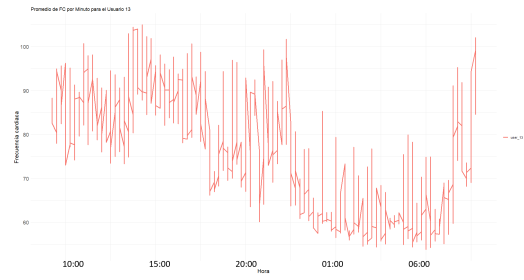


Figura 26: Frecuencia cardiaca por hora para el user_13 con $CI=0.81$.

En el caso del usuario 10, se observan valores atípicos tanto en los máximos como en los mínimos de la frecuencia cardiaca, así como valores ausentes al final del experimento. Estos resultados sugieren un ritmo circadiano menos robusto.

En contraste, el usuario 13 muestra una variación en la frecuencia cardiaca que puede explicarse por el comportamiento esperado del ritmo circadiano: se observan aumentos durante las horas de vigilia y disminuciones tanto después de las comidas como durante la noche. Estos patrones reflejan un comportamiento circadiano que se alinea con las expectativas normales del ritmo biológico de 24 horas.

Finalmente, investigamos la relación entre la variabilidad intradiaria (IV) y las características de los usuarios. La matriz de correlación entre IV y otras variables se muestra en la Figura 27.

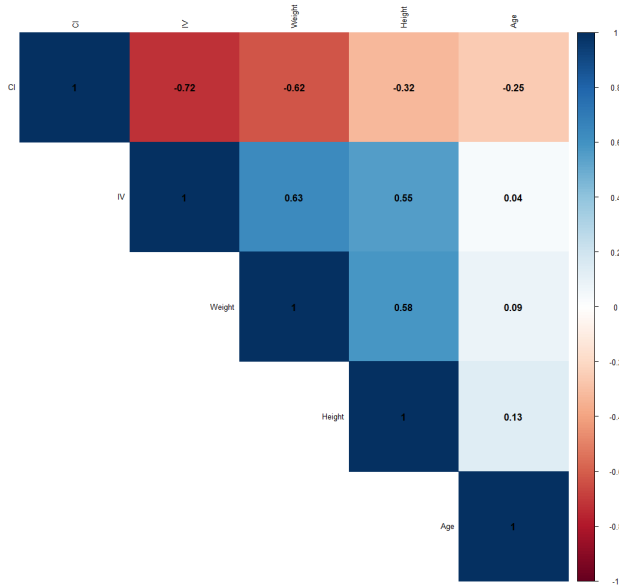


Figura 27: Correlación entre características, CI e IV.

La correlación negativa observada entre los índices CI e IV, con un valor de -0.72, sugiere que a medida que aumenta el CI, disminuye el IV. Esto es consistente con la expectativa de que un sistema circadiano bien desarrollado, representado por un CI elevado, muestra una mayor estabilidad diaria, lo que se traduce en una menor frecuencia de transiciones entre periodos de actividad y reposo, y, por lo tanto, un menor IV.

La correlación observada entre el peso y los índices CI e IV presenta el mayor valor en términos absolutos entre las variables analizadas, lo que sugiere una posible relación entre el peso corporal y el sistema circadiano. Es relevante señalar que dicha correlación es prácticamente idéntica para ambos índices, lo cual refuerza la hipótesis de que el peso podría influir tanto en el CI como en el IV.

Dado que también se dispone de los datos de altura, y considerando que el peso por sí solo no constituye un indicador exhaustivo del estado de salud de un individuo, se procedió a calcular el Índice de Masa Corporal (IMC) mediante la fórmula:

$$\text{IMC} = \frac{\text{Peso}}{(\text{Altura})^2}$$

El IMC es una métrica ampliamente aceptada que vincula el peso con la altura de

la persona. Valores de IMC superiores a 25 se consideran indicativos de sobrepeso, mientras que un IMC mayor a 30 corresponde a la categoría de obesidad. Por otro lado, valores de IMC inferiores a 18.5 pueden indicar desnutrición o bajo peso, lo cual también conlleva riesgos para la salud, como debilidad general y problemas nutricionales. Tras calcular la correlación entre el CI e IV, y el IMC, se observó una reducción de la correlación hasta $-0,56$ y $0,42$ respectivamente. La Figura 28 representa la relación entre las variables IMC y CI.

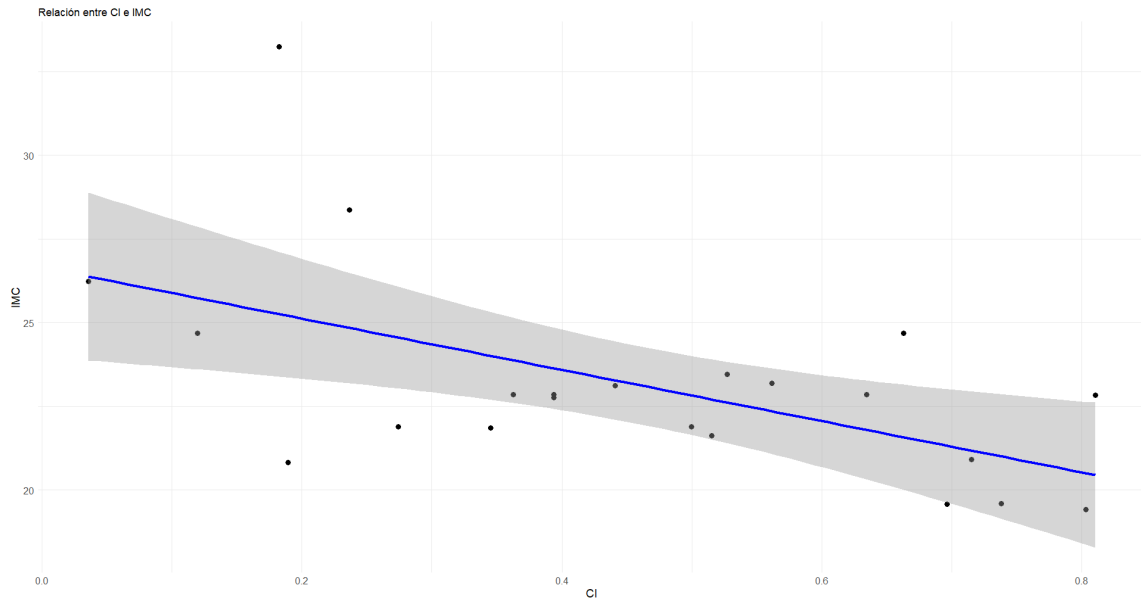


Figura 28: Relación entre IMC y CI.

La tendencia negativa observada entre nuestros usuarios sugiere que aquellos con un peso dentro del rango considerado normal presentan un ritmo circadiano de mayor calidad. Para confirmar esta relación, sería necesario realizar el experimento con un mayor número de participantes.

5 Conclusiones

El modelo FMM ha demostrado un rendimiento superior en comparación con el modelo cosinor, especialmente en casos con comportamientos no sinusoidales. Esta parametrización ofrece una mayor flexibilidad para modelar dinámicas complejas que los modelos sinusoidales convencionales no pueden capturar de manera adecuada.

El uso del ACPF ha demostrado ser útil para identificar los momentos de mayor variabilidad en los datos, así como para detectar tendencias mediante la aproximación de los datos utilizando las puntuaciones, las autofunciones y la media del proceso.

Los índices no paramétricos, como la amplitud relativa, el índice de circadianidad y la variabilidad intradiaria, también han sido efectivos para relacionarse con otras variables, como las características del usuario o los biomarcadores.

Sin embargo, no hemos encontrado ninguna relación entre los datos subjetivos obtenidos a través de cuestionarios y variables que recogen datos del sueño, como el tiempo en la cama o el número de despertares.

Hemos presentado distintos métodos paramétricos y no paramétricos y aplicado estos mismos al conjunto de datos MMASH. Estas técnicas pueden ser utilizadas siguiendo el mismo esquema para otros conjuntos de datos. Además, hemos diseñado una nueva medida que combina la frecuencia cardíaca, el movimiento y los cambios de posición del cuerpo, reduciendo así el posible ruido en cada una de estas variables.

El conjunto de datos MMASH solo tiene datos de 24 horas por lo que los índices relacionados con la estabilidad del ritmo circadiano no han podido ser calculados. Tampoco disponemos de valores de temperatura que podríamos haber utilizado para calcular índices no paramétricos relacionados con la sincronización tanto interna como ambiental.

A la hora de tratar con señales hemos utilizado principalmente el modelo FMM y el ACPF, habiéndose utilizado el primero para agrupar los distintos usuarios. Una posible línea de trabajo sería haber utilizado componentes funcionales principales multivariantes y, en vez de aplicar ACPF a la nueva medida haberlo aplicado a un conjunto de variables [28].

Otra forma de trabajar con este tipo de datos, sobre todo si queremos hacer clustering, es usar técnicas de deep learning como las CNNs, LSTM o GRU para series temporales [29].

A Teoremas del análisis de datos funcionales

Un núcleo es una función continua y simétrica

$$K : [a, b] \times [a, b] \rightarrow \mathbb{R}$$

donde simetría significa que

$$K(x, y) = K(y, x)$$

para todos $x, y \in [a, b]$.

Se dice que K es un núcleo definido positivo si, y solo si,

$$\sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) c_i c_j \geq 0$$

para cualquier secuencia finita de puntos $x_1, \dots, x_n \in [a, b]$ y cualquier elección de números reales c_1, \dots, c_n .

El núcleo K está asociado a un operador lineal (en particular, un operador integral de Hilbert–Schmidt) que actúa sobre funciones, definido por la integral

$$[T_K \varphi](x) = \int_a^b K(x, s) \varphi(s) ds.$$

Donde φ pertenece al espacio $L^2[a, b]$, que es el conjunto de funciones de valores reales cuadráticamente integrables. Dado que T_K es un operador lineal, podemos hablar de sus autovalores y autofunciones.

Teorema 2. Teorema de Mercer

Sea K un núcleo continuo, simétrico y definido positivo. Entonces, existe una base ortonormal $\{e_i\}_i$ de $L^2[a, b]$ compuesta por autofunciones del operador asociado al núcleo integral T_K tal que la secuencia correspondiente de autovalores $\{\lambda_i\}_i$ es no negativa. Las autofunciones correspondientes a los autovalores no nulos son continuas en $[a, b]$, y K tiene la siguiente representación:

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j e_j(s) e_j(t)$$

donde la convergencia es absoluta y uniforme.

La situación en este enfoque es la siguiente:

Consideraremos un proceso estocástico $X_t = X(t)$ definido sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ y un intervalo cerrado $[a, b]$. Además es de cuadrado-integrable, tiene media cero y con función de covarianza $K_X(s, t)$.

Resumiendo:

- $\forall t \in [a, b], \quad X_t \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, es decir, $\mathbb{E}[X_t^2] < \infty$,
- $\forall t \in [a, b], \quad \mathbb{E}[X_t] = 0$,
- $\forall t, s \in [a, b], \quad K_X(s, t) = \mathbb{E}[X_s X_t]$.

La condición $\mathbb{E}[X_t^2] < \infty$ es equivalente a que $K_X(s, t)$ sea finita $\forall s, t \in [a, b]$. Llamamos T_{K_X} al operador lineal asociado al núcleo integral K_X y está definido de la siguiente manera:

$$T_{K_X} : L^2([a, b]) \rightarrow L^2([a, b]) : \quad f \mapsto T_{K_X} f = \int_a^b K_X(s, \cdot) f(s) ds.$$

Dado que T_{K_X} es un operador lineal, tiene sentido hablar de sus valores propios λ_k y autofunciones e_k , que se encuentran resolviendo la ecuación integral:

$$\int_a^b K_X(s, t) e_k(s) ds = \lambda_k e_k(t).$$

Teorema 3. Teorema de Karhunen–Loève

Sea X_t un proceso estocástico cuadrado-integrable con media cero definido sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ en $[a, b]$, con función de covarianza continua $K_X(s, t)$.

Entonces $K_X(s, t)$ es un núcleo al que podemos aplicar el Teorema de Mercer y, si e_k es una base ortonormal en $L^2([a, b])$ formada por las autofunciones de T_{K_X} con valores propios respectivos λ_k , X_t admite la siguiente representación:

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t),$$

donde la convergencia es en L^2 , uniforme en t y

$$Z_k = \int_a^b X_t e_k(t) dt.$$

Además, las variables aleatorias Z_k tienen media cero, son no correlacionadas y tienen varianza λ_k :

$$\mathbf{E}[Z_k] = 0, \quad \forall k \in \mathbb{N} \quad \text{y} \quad \mathbf{E}[Z_i Z_j] = \delta_{ij} \lambda_j, \quad \forall i, j \in \mathbb{N}.$$

Notemos que, por generalizaciones del teorema de Mercer, podemos reemplazar el intervalo $[a, b]$ por otros espacios compactos C y la medida de Lebesgue en $[a, b]$ por una medida de Borel cuyo soporte sea C .

B Tablas complementarias

Variable	mean	sd	min	p25	p50	p75	max
Latency	1.5	1.71	0	0	0.5	3	4
Efficiency	83.9	6.75	73.5	77.2	85.2	89.1	94.2
TMB	374.	96.2	165	330.	368.	431	630
TST.	313	84.3	144	254.	326	343.	578
WASO.	59.8	30.5	17	40	52.5	77.8	118
Number of Awakenings	19.3	9.78	4	12.2	18.5	21	44
Average.Awakening Length	3.56	2.39	1.33	2.08	2.80	4.25	12.2
MI	13.5	4.38	6.73	9.16	13.2	17.4	20.7
IFs	10.3	9.22	0	0	9.76	15.7	28.1
SFI	23.8	11.5	6.73	16.2	22.1	30.8	45.5

Tabla 14: Estadísticos resumen de las variables del archivo sleep.csv.

Usuario	mean	sd	min	p25	p50	p75	max
user_1	0.7123	0.1858	0.251	0.597	0.679	0.813	27.759
user_2	0.8276	0.4784	0.250	0.697	0.788	0.881	61.688
user_3	0.7779	0.2018	0.055	0.667	0.748	0.877	11.099
user_4	0.9958	0.3695	0.251	0.737	0.873	1.121	21.953
user_5	0.8073	0.2531	0.163	0.698	0.807	0.897	41.942
user_6	0.7040	0.2017	0.250	0.602	0.675	0.805	22.367
user_7	0.7925	0.1665	0.250	0.667	0.758	0.934	4.729
user_8	0.8318	0.2310	0.250	0.728	0.820	0.928	16.857
user_9	0.8217	0.2548	0.251	0.684	0.882	1.003	17.203
user_10	0.8291	0.2792	0.251	0.720	0.817	0.922	18.514
user_11	0.8570	0.1633	0.253	0.744	0.844	0.967	15.220
user_12	0.8156	0.1781	0.252	0.711	0.805	0.923	6.986
user_13	0.7989	0.1820	0.251	0.667	0.761	0.939	20.414
user_14	0.8340	0.1498	0.282	0.740	0.830	0.943	1.427
user_15	0.6710	0.3019	0.250	0.541	0.642	0.776	24.496
user_16	0.8602	0.2183	0.250	0.726	0.813	0.939	15.441
user_17	0.7106	0.3565	0.250	0.577	0.662	0.798	61.561
user_18	0.8545	0.1849	0.250	0.764	0.845	0.945	12.413
user_19	0.8198	0.2028	0.250	0.656	0.768	1.007	8.393
user_20	0.8644	0.2203	0.250	0.729	0.823	0.979	13.756
user_21	0.7922	0.1526	0.253	0.688	0.794	0.898	5.112
user_22	0.9220	0.1804	0.252	0.808	0.886	1.004	14.503

Tabla 15: Estadísticos resumen del IBI para cada usuario.

	mean	sd	min	p25	p50	p75	max
MEQ	50.32	7.03	38.00	45.25	49.50	55.00	64.00

Tabla 16: Estadísticos resumen para el cuestionario MEQ.

Noche (MEQ \leq 41)	Inter (42 \leq MEQ \leq 58)	Mañana (MEQ \geq 59)
user_10 user_11	user_1 user_2 user_5 user_6 user_7 user_9 user_12 user_13 user_15 user_16 user_17 user_18 user_19 user_20 user_21 user_22	user_3 user_4 user_8 user_14

Tabla 17: Clasificación de usuarios basada en las puntuaciones de MEQ.

Buena Calidad Sueño (PSQI \leq 5)	Mala Calidad Sueño (PSQI \geq 5)
user_1 user_4 user_8 user_10 user_12 user_13 user_14 user_15 user_16 user_17 user_18 user_19 user_20 user_21 user_22	user_2 user_3 user_5 user_6 user_7 user_9 user_11

Tabla 18: Clasificación de usuarios basada en las puntuaciones de Pittsburgh.

Usuario	mean	sd	min	p25	p50	p75	max
user_1	84.7	16.9	32	70	86	97	251
user_2	74.5	16.9	10	67	75	82	233
user_3	77.5	18.1	3	62	78	85	194
user_4	71.9	11.8	44	63	72	81	189
user_5	74.9	12.3	3	66	73	82	228
user_6	86.3	15.6	46	72	87	97	224
user_7	76.2	15.3	47	62	76	87	223
user_10	73.5	11.7	8	66	72	80	218
user_11	70.4	11.1	16	61	70	78	213
user_12	73.9	16.9	49	61	73	81	166
user_13	75.4	14.5	15	61	75	87	145
user_14	72.2	13.2	47	62	70	78	207
user_15	91.5	28.5	12	70	89	102	240
user_16	69.8	15.3	10	55	71	80	224
user_17	85.9	21.4	20	68	87	99	228
user_18	70.5	11.5	16	63	70	75	228
user_19	73.6	16.7	3	57	72	87	152
user_20	70.0	14.7	13	57	70	80	233
user_21	76.1	14.6	51	65	74	83	197
user_22	65.3	11.1	42	57	67	72	131

Tabla 19: Estadísticos resumen de la frecuencia cardiaca para cada usuario.

Referencias

- [1] Z. Shan, H. Ma, M. Xie, P. Yan, Y. Guo, W. Bao, Y. Rong, C. L. Jackson, F. B. Hu, and L. Liu. Sleep duration and risk of type 2 diabetes: a meta-analysis of prospective studies. *Diabetes care*, 38(3):529–537, 2015.
- [2] Q. Wang, B. Xi, M. Liu, Y. Zhang, and M. Fu. Short sleep duration is associated with hypertension risk among adults: a systematic review and meta-analysis. *Hypertension Research*, 35(10):1012–1018, 2012.
- [3] K. G. Baron and K. J. Reid. Circadian misalignment and health. *International review of psychiatry*, 26(2):139–154, 2014.
- [4] A. Rossi, E. Da Pozzo, D. Menicagli, C. Tremolanti, C. Priami, A. Sîrbu, and D. Morelli. A public dataset of 24-h multi-levels psycho-physiological responses in young healthy adults. *Data*, 5(4):91, 2020.
- [5] Françoise Lévi, Armagan Okyar, Sylvain Dulong, Pascal F. Innominato, and Jean Clairambault. Circadian timing in cancer treatments. *Annual Review of Pharmacology and Toxicology*, 50(1):377–421, 2010.
- [6] C. W. Pickering, K. A. Foster, and P. G. H. Khatami. Non-parametric methods for analyzing circadian rhythms. *Chronobiology International*, 30(5):542–553, 2013.
- [7] E. Ortiz-Tudela, A. Martinez-Nicolas, J. Albares, F. Segarra, M. Campos, E. Estivill, and J. A. Madrid. Ambulatory circadian monitoring (acm) based on thermometry, motor activity and body position (tap): a comparison with polysomnography. *Physiology & behavior*, 126:30–38, 2014.
- [8] A. Martinez-Nicolas, M. J. Martinez-Madrid, P. F. Almada-Pagan, M. A. Bonmati-Carrion, J. A. Madrid, and M. A. Rol. Assessing chronotypes by ambulatory circadian monitoring. *Frontiers in physiology*, 10:1396, 2019.
- [9] Fred Shaffer and Julian P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5:258, 2017.
- [10] Task Force of the European Society of Cardiology Electrophysiology and the North American Society of Pacing and. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065, 1996.

- [11] K. Li, H. Rüdiger, and T. Ziemssen. Spectral analysis of heart rate variability: time window matters. *Frontiers in Neurology*, 10:545, 2019.
- [12] Sebastian Wallot, Riccardo Fusaroli, Knud Tylén, and E. M. Jegindø. Using complexity metrics with rr intervals and bpm heart rate measures. *Frontiers in Physiology*, 4:211, 2013.
- [13] H. M. van Goor, K. van Loon, M. J. Breteler, C. J. Kalkman, and K. A. Kaasjager. Circadian patterns of heart rate, respiratory rate and skin temperature in hospitalized covid-19 patients. *PLoS One*, 17(7):e0268065, 2022.
- [14] C. Rueda, Y. Larriba, and S. D. Peddada. Frequency modulated möbius model accurately predicts rhythmic signals in biological and physical sciences. *Scientific reports*, 9(1):18701, 2019.
- [15] I. Fernández, A. Rodríguez-Collado, Y. Larriba, A. Lamela, C. Canedo, and C. Rueda. fmm: An r package for modeling rhythmic patterns in oscillatory systems. 2021. Available at: <https://arxiv.org/abs/2105.10168>.
- [16] Y. Larriba, C. Rueda, M. A. Fernández, and S. D. Peddada. Order restricted inference in chronobiology. *Statistics in Medicine*, 39(3):265–278, 2020.
- [17] G. Cornelissen. Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling*, 11:1–24, 2014.
- [18] D. Fekedulegn, M. E. Andrew, M. Shi, J. M. Violanti, S. Knox, and K. E. Innes. Actigraphy-based assessment of sleep parameters. *Annals of Work Exposures and Health*, 64(4):350–367, 2020.
- [19] J. Vicente-Martínez, P. F. Almada-Pagan, A. Martínez-Nicolas, J. A. Madrid, M. A. Rol, and M. Á. Bonmatí-Carrión. *Integral analysis of circadian rhythms*. Springer International Publishing, Cham, 2023.
- [20] E. Ortiz-Tudela, A. Martínez-Nicolas, M. Campos, M. Á. Rol, and J. A. Madrid. A new integrated variable based on thermometry, actimetry and body position (tap) to evaluate circadian system status in humans. *PLoS computational biology*, 6(11):e1000996, 2010.
- [21] J. Gertheiss, D. Rügamer, B. X. Liew, and S. Greven. Functional data analysis: An introduction and recent developments. *arXiv preprint*, arXiv:2312.05523, 2023.

- [22] S. Luczak, R. Grepl, and M. Bodnicki. Selection of mems accelerometers for tilt measurements. *Journal of Sensors*, 2017(1):9796146, 2017.
- [23] A. Rossi, E. Da Pozzo, D. Menicagli, C. Tremolanti, C. Priami, A. Sirbu, and D. Morelli. A public dataset of 24-h multi-levels psycho-physiological responses in young healthy adults. *Data*, 5(4):91, 2020.
- [24] V. Natale, M. J. Esposito, M. Martoni, and M. Fabbri. Validity of the reduced version of the morningness-eveningness questionnaire. *Sleep and Biological Rhythms*, 4(1):72–74, 2006.
- [25] M. D. Manzar, A. S. BaHamam, U. A. Hameed, D. W. Spence, S. R. Pandi-Perumal, A. Moscovitch, and D. L. Streiner. Dimensionality of the pittsburgh sleep quality index: a systematic review. *Health and quality of life outcomes*, 16:1–22, 2018.
- [26] Fionn Murtagh and Pedro Contreras. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.
- [27] Y. Januzaj, E. Beqiri, and A. Luma. Determining the optimal number of clusters using silhouette score as a data mining technique. *International Journal of Online & Biomedical Engineering*, 19(4), 2023. Available online.
- [28] J. M. Chiou, Y. T. Chen, and Y. F. Yang. Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 24:1571–1596, 2014.
- [29] B. Lafabregue, J. Weber, P. Gañarski, and G. Forestier. End-to-end deep representation learning for time series clustering: a comparative study. *Data Mining and Knowledge Discovery*, 36(1):29–81, 2022.