
Universidad de Valladolid

FACULTAD DE CIENCIAS

# Estudio de factores influyentes en la ritmicidad de expresiones de genes a partir de datos post mortem

Autora:
**Carmen Pazos Páramo**

Tutores:
**Yolanda Larriba González**
**Miguel Alejandro Fernández Temprano**

Trabajo de fin de grado del
Grado en Estadística

Universidad de Valladolid
Marzo 2022

## RESUMEN

Los genes circadianos, generalmente vinculados a los ciclos de luz-oscuridad, resultan vitales en la gran mayoría de los procesos cerebrales, al regular las funciones biológicas básicas de nuestro organismo.

Estos genes presentan un patrón rítmico que se repite cada 24 horas. El análisis de esta ritmicidad resulta muy relevante para identificar posibles patologías. Sin embargo, obtener este tipo de datos puede poner en riesgo la salud de los pacientes y, por ende, es habitual trabajar con muestras post-mortem, en las que generalmente se desconoce el momento del día en el que se toman las muestras.

Entre las contribuciones de este proyecto se encuentra el estudio de factores como el sexo, la edad o la causa de la muerte. Para lograr este objetivo, se utilizará metodología estadística para estimar los órdenes entre los instantes de tiempo que recuperan la expresión rítmica propia de los genes circadianos; modelos para el análisis de procesos oscilatorios con datos direccionales, y diferentes estimadores de la asociación circular. Los resultados de este proyecto suponen un estudio crítico de los efectos que cada uno de los factores analizados en la expresión rítmica de los genes.



## ABSTRACT

Circadian genes are important in nearly all processes in the brain, related to the day-night cycle, governing basic biological functions.

These genes display rhythmic temporal patterns, the pattern analysis is the key to identify pathologies. However, obtaining this type of data can put the health of patients at risk and, therefore, it is common to work with post-mortem samples, in which the time of day that the samples are taken is generally unknown.

The contribution of this work is to study the changes in rhythms that comes with factors like gender, cause of death or aging. For this porpuse, a statistical methodology is used in order to estimate the orders between the instants of time that recover the rhythmic expression of circadian genes, in conjunction with models for the analysis of oscillatory processes with directional data and different estimators of the circular association. The results of this project present a critical study of the effects that each of the factors analyzed in the rhythmic expression of genes.

# CONTENIDO

# 1.INTRODUCCIÓN

El ser humano, al igual que el resto de seres vivos, posee un “reloj interno” que se adapta a las diferentes horas y momentos del día. A estos “relojes internos” se les conoce como ritmos circadianos y, aunque se empezaron a investigar a inicios del siglo XIX, no fue hasta la década de los 80 cuando se descubrió el mecanismo molecular que lo regulaba, produciendo grandes avances en la biomedicina y la cronobiología, disciplina que estudia los fenómenos cíclicos o ritmos biológicos en los seres vivos, (Borja, 2019) y (Cano, 2018).

Los ritmos circadianos son las fluctuaciones periódicas que se producen a lo largo del día en nuestro organismo, estos ciclos se repiten constantemente en periodos de 24 horas, causados por la sintonía con el medio ambiente y los estímulos lumínicos. Los lapsos luz-oscuridad son captados por nuestro organismo y procesados por el núcleo supraquiasmático, una región del cerebro que funciona como un “reloj central”, emitiendo órdenes que armonizan el resto de nuestro cuerpo (Lorsch, 2021).

Los comportamientos o hábitos que provocan una alteración relevante en los ritmos circadianos pueden causar lo que se denomina como cronodisrupción. Numerosos estudios prueban una asociación elevada entre la cronodisrupción y un incremento en alteraciones del sueño, enfermedades cardiovasculares, deterioros cognitivos, envejecimiento e, incluso, algunos tipos de cáncer (Borja, 2019).

Los genes que regulan estos ritmos reciben el nombre genes circadianos. Asimismo, su expresión genética es el proceso por el cual todos los organismos transforman la información codificada por los ácidos nucleicos en proteínas. Siendo, finalmente las proteínas las encargadas de regular y controlar las funciones del organismo. Este proceso está estrictamente regulado y permite que una célula responda a los cambios en su entorno, controlando la síntesis de las proteínas y estas, a su vez, los diferentes procesos biológicos (Griffiths & Lewontin, 2004). Por consiguiente, la expresión de los genes circadianos es un indicador del estado de las células y puede asociarse, entre otras, con la aparición de enfermedades neurogenerativas, depresión o trastornos metabólicos (Liu et al., 2017).

El análisis de los patrones y ritmos circadianos puede suponer un gran avance en la detección temprana de algunas enfermedades. No obstante, aún no existe ningún tipo de procedimiento médico para la extracción de estos genes sin poner en riesgo la salud o hacer peligrar la vida del paciente. Por ello, es habitual que el análisis de los ritmos circadianos se efectúe a partir de muestras de personas ya fallecidas, en adelante muestras post-mortem. En consecuencia, el momento exacto de la muerte puede ser desconocido o impreciso, como es el caso por ejemplo de cuando la defunción se debe a un infarto, puesto que las funciones biológicas y, por extensión, la expresión de los genes, continúan su curso tras la muerte clínica (Piacente, 2021).

Por consiguiente, en la práctica es habitual disponer de muestras post-mortem de expresión de gen, bien desordenadas, o bien con una imprecisión en el momento del día en que estas fueron tomadas. Es decir, como etapa previa es necesaria una estimación del orden temporal de los instantes de tiempo de las muestras para poder analizar posteriormente, una vez que las muestras estén ordenadas, los ritmos circadianos. El método más extendido en la literatura es la ordenación por el momento de fallecimiento estimado o ToD, sin embargo, por los motivos expuestos previamente puede producir estimaciones imprecisas, por lo que resulta necesario implementar nuevas técnicas, tales como Oscope (Leng et al., 2015), CYCLOPS (Anafi et al., 2017) o la metodología basada en Inferencia con Restricciones de Orden, en adelante ORI, empleada para obtener la ordenación óptima de la expresión de estos genes mediante permutaciones entre los instantes de tiempo y que será la utilizada en este trabajo.

Una vez conseguida la ordenación de los datos, se emplean diferentes modelos para ajustar la señal rítmica de los genes. Cabe destacar que la señal rítmica subyacente en procesos oscilatorios (up-down-up), como el caso del patrón de expresión de genes circadianos, puede formularse matemáticamente como una *señal circular.* El carácter periódico innato de estas señales hace que esta formulación sea equivalente tanto en el espacio euclídeo como en el espacio circular, y que por ende se utilicen modelos para datos direccionales. En este proyecto se consideran dos modelos paramétricos: Cosinor (Cornelissen, 2014), el más clásico y más empleado en la literatura y FMM (Rueda et al., 2019) que permite, a diferencia del Cosinor, el ajuste de patrones rítmicos asimétricos. Por otro lado, también se emplea un modelo no paramétrico (Larriba et al., 2020). Todos estos métodos mencionados permiten valorar el desempeño de la ordenación ORI.

Además del ajuste de la señal, un marcador clave en la biología circadiana es el momento de máxima expresión del gen o *peak*, es decir, el instante temporal en el que el gen alcanza su máximo valor de expresión y que está asociado con el instante temporal (o momento del día) en el que el gen lleva a cabo su función biológica.

Tal y como se ha mencionado antes, múltiples causas pueden estar asociadas con la ruptura de los ritmos circadianos. Por ende, uno de los objetivos principales de este proyecto consiste comparar los patrones rítmicos entre hombres y mujeres o motivos de fallecimiento, atributos que hasta donde se sabe, no se han estudiado hasta el momento en la literatura. Para ello, se analizará la asociación de los momentos de máxima expresión de un conjunto bien conocido en la literatura de genes rítmicos, en adelante *genes core*, entre las distintas categorías de variables como sexo o causa de muerte. Nótese que los *peaks* son datos direccionales, es decir toman valores en el círculo unidad, y en consecuencia, para poder realizar este estudio, se ha requerido de dos coeficientes de correlación circular: el coeficiente de Jammalamadaka (Jammalamadaka & Sarma, 1988) y de Fisher-Lee (Fisher & Lee, 1995). Estas técnicas han permitido analizar y cuantificar los cambios en la ritmicidad

de los genes originados por los factores anteriormente mencionados, objeto de estudio de este proyecto.

## 1.1 Objetivos

El objetivo fundamental del proyecto es el estudio de la posible influencia de variables categóricas dicotómicas en los momentos de máxima expresión de genes circadianos, mediante la estimación del grado de asociación entre los momentos de máxima expresión obtenidos en cada categoría de las variables dicotómicas (origen, sexo y causa de muerte) a través de coeficientes de correlación circulares.

Para llegar a este objetivo se han tenido que cubrir las siguientes etapas previas que han supuesto un aprendizaje de conceptos no desarrollados en las asignaturas del grado:

- Estimación, mediante la metodología ORI, del orden temporal de datos de expresiones de genes cuyo orden es desconocido al tratarse una muestra post-mortem. Esta etapa supone el estudio previo de la metodología de la inferencia con restricciones de orden y el manejo del problema del viajante (TSP).
- Ajuste de las expresiones de los genes mediante los modelos de señal oscilatoria para cada variable dicotómica. Los modelos considerados en esta etapa (FMM, Cosinor y no paramétrico) no son modelos para datos euclídeos como los considerados en el grado sino para datos direccionales lo que ha supuesto la inmersión en estos tipos de problema que no han sido tratados en el grado.
- Estudio de los diferentes coeficientes de correlación para datos circulares propuestos en la literatura (Jammalamadaka, Fisher-Lee) para valorar su utilidad en los problemas que aquí se presentan.

## 1.2 Estructura del documento

La memoria de este proyecto se compone de los siguientes epígrafes:

- Metodología: se explica el método para ordenar la expresión de los genes, los modelos para validarlo y las medidas de asociación entre datos circulares.
- Datos: se describe los datos empleados en este proyecto, así como las transformaciones empleadas para su tratamiento.
- Resultados: se valora los ajustes obtenidos de cada gen y variable para cada modelo empleado. Asimismo, se estudia la asociación entre las variables dicotómicas.

- Conclusiones: se resumen los resultados obtenidos y los posibles trabajos futuros.
- Anexos: se incluye el código R empleado.

## 1.3 Asignaturas relacionadas

En esta sección se muestra la relación de técnicas empleadas en este proyecto y el aprendizaje de estas en diferentes asignaturas del grado:

- Computación estadística: en esta asignatura se asientan las bases sobre el lenguaje de programación R, usado en este trabajo.
- Probabilidad, Inferencia y Análisis de datos categóricos: en ellas se estudia el fundamento y las bases de los coeficientes de correlación, intervalos de confianza y métricas utilizados en este proyecto.
- Modelo de Investigación Operativa y Algoritmos de computación: donde se dan soluciones al TSP (Problema del viajante).
- Modelos lineales: en esta asignatura se estudia el fundamento y base para la comprensión de los modelos estadísticos.

## 1.4 Ampliación de la materia

Los contenidos ampliados respecto a los estudiados en el grado son los siguientes:

- Método ORI: metodología propuesta para la estimación del orden temporal.
- Modelos FMM, Cosinor y no paramétrico: modelos de señal oscilatoria empleados para la validación del orden temporal.
- Análisis de datos direccionales.
- Coeficientes de correlación de Jammadamalaka y Fisher-Lee: usados para estudiar la asociación entre las variables.

# 2. METODOLOGÍA

En esta sección se definen los conceptos y técnicas elementales para estudiar la influencia de variables tales como el sexo, la edad, la causa de la muerte o el propio origen de los datos en la expresión del gen.

En primer lugar, se introduce el concepto de señal up-down-up, y su formulación equivalente entre los espacios euclídeo y circular. La periodicidad intrínseca de las señales up-down-up, utilizadas para el análisis de procesos oscilatorios, propicia que estas señales estén ligadas al espacio circular, mientras que su representación en el espacio euclídeo resulta más inmediata para estudiar los patrones de expresión del gen, al ser este el espacio donde se obtienen los datos.

A continuación, se explican los métodos empleados en este proyecto para la ordenación y validación de los ritmos circadianos. Estas técnicas son, respectivamente, la metodología ORI y los modelos Cosinor, FMM y no paramétrico.

Los epígrafes finales de esta sección se centran en la explicación de técnicas específicas para el estudio estadístico de los datos direccionales. En particular, se presentan los estimadores de correlación circular de Jammalamadaka y de Fisher-Lee que se utilizan en este trabajo.

## 2.1 Señales circulares en procesos oscilatorios: Señal up-down-up

Esta sección propone el uso de señales *circulares* para modelar procesos oscilatorios. El carácter periódico y rítmico de estos fenómenos, hace que se puedan mapear como procesos *circulares,* tanto en el espacio euclídeo como en el espacio circular (Larriba et al., 2020).

En el espacio euclídeo, las señales que rigen el comportamiento de estos procesos siguen un patrón rítmico de expresión up-down-up que se puede describir mediante desigualdades matemáticas que, establecen restricciones de orden entre los valores de la señal, (véase Definición 1).

Se va a denotar por $X_{ij}$ la observación recogida para el gen $j$ en el instante temporal $t_i$, con $i = 1, \dots, n$. Se denotará también como $\boldsymbol{X_j} = (X_{1j}, \dots, X_{nj})'$ al vector de datos para el gen $j$, con $j = 1, \dots, J$, a lo largo de todos los $n$ instantes de tiempo.

El gen $j$ sigue un modelo de señal circular, si:

$$\boldsymbol{X_j} = \boldsymbol{\mu_j} + \boldsymbol{\varepsilon_j}, j = 1 \dots J. \quad (1)$$

Donde $\boldsymbol{\mu_j}$ es una señal up-down-up que describe un patrón monótonamente creciente hasta el instante $t_U$ - monótonamente decreciente hasta el instante $t_L$, y

que a partir de ese instante vuelve a crecer monótonamente. Y donde $\boldsymbol{\varepsilon_j}$ denota el error aleatorio para el que se asume una distribución normal con vector de medias 0 y matriz de varianzas diagonal.

A continuación, se define matemáticamente el concepto de señal circular up-down-up en el espacio euclídeo y su equivalente en el espacio circular. Las siguientes definiciones se corresponden para el caso $U < L$ y se pueden dar de forma análoga cuando $L < U$:

Definición 1: Una señal $\boldsymbol{\mu}$ en el espacio euclídeo es up-down-up si y solo si:

$$\boldsymbol{\mu} \in C = \bigcup_{L,U} C_{LU} \tag{2}$$

donde $L = argmin_{1..n}\mu_i$, $U = argmax_{1..n}\mu_i$ y $C_{LU} = \{\boldsymbol{\mu} \in \mathbb{R}^n : \mu_1 \leq \cdots \leq \mu_U \geq \cdots \geq \mu_L \leq \cdots \leq \mu_L \leq \mu_1\}$.

Nótese que una señal circular up-down-up $\boldsymbol{\mu}$ definida en el espacio euclídeo, induce una señal circular up-down-up $\boldsymbol{\phi}$ en el espacio circular ($\boldsymbol{\phi} \in [0,2\pi]^n$), tal que $\phi_1 \preceq \cdots \preceq \phi_U \preceq \cdots \preceq \phi \preceq \cdots \preceq \phi_n \preceq \phi_1$, es decir $\boldsymbol{\phi}$ sigue el orden circular $\boldsymbol{o} = (o_1, \ldots, o_{U,}, \ldots, o_L, \ldots, o_n)$. Como veremos a continuación, el recíproco también es cierto.

Definición 2: Una señal $\boldsymbol{\phi}$ en el espacio circular es up-down-up si y solo sí:

$$\boldsymbol{\phi} \in \mathrm{C}_{\boldsymbol{o}} = \{\boldsymbol{\phi} \in [0,2\pi]^n : \phi_1 \preceq \cdots \preceq \phi_n \preceq \phi_1\} \tag{3}$$

es decir $\boldsymbol{\phi}$ sigue el orden circular $\boldsymbol{o}$. La notación " $\preceq$ " se lee como *es seguido por*.

Finalmente, la siguiente formula representa la equivalencia entre una señal up-down-up en el espacio euclídeo y en el espacio el circular:

$$\phi_i = T_{LU}(\mu_i) = \begin{cases} \arcsin(\mu_i), & si\ i \in \{1, \ldots, U\} \cup \{L, \ldots, n\} \\ \frac{\pi}{2} - \arcsin(\mu_i), & en\ el\ resto\ de\ casos \end{cases} \tag{4}$$

La Figura 1 es una representación de la equivalencia dada. En la parte izquierda se puede observar una la señal circular $\boldsymbol{\mu}$ en el espacio euclídeo con el respectivo patrón up-down-up, mientras a la derecha se muestra una señal circular $\boldsymbol{\phi}$ siguiendo el orden circular $\boldsymbol{o}$.

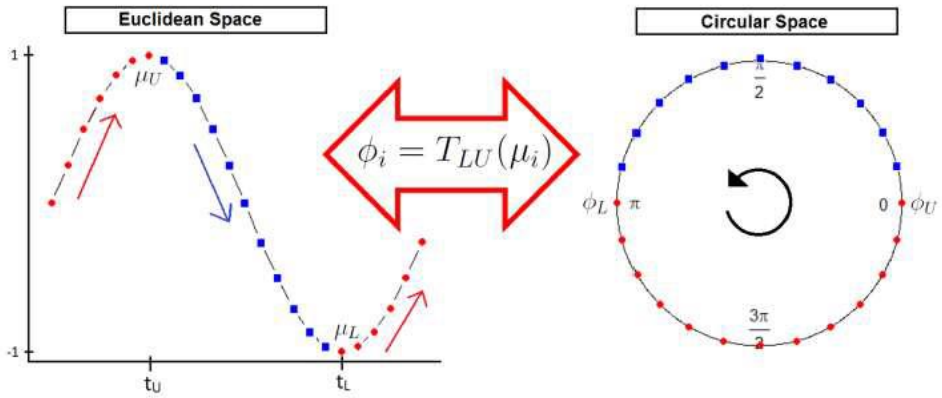


Figura 1: Imagen que representa la equivalencia entre una señal up-down-up en el espacio euclídeo y en el espacio circular. Fuente: Larriba, Rueda, Fernández, & Peddada, 2019.

## 2.2 Métodos de estimación del orden temporal:

El análisis de procesos oscilatorios requiere que el orden entre los instantes de tiempo sea conocido, sin embargo, en la práctica esto no es siempre posible.

En esta sección se describe con más detalle la metodología empleada para buscar el orden de los índices (instantes de tiempo), cuando este es desconocido y ha de estimarse, de modo que los datos recogidos se adapten lo mejor posible al modelo up-down-up.

Entre los métodos existentes cabe destacar: el algoritmo Oscope (Leng et al., 2015) que permite recuperar el ciclo celular en datos RNA; el algoritmo CYCLOPS (Anafi et al., 2017), basado en redes neuronales, que utiliza información a priori sobre los genes y funciona como una *black-box;* y la metodología ORI, basada en inferencias con restricciones de orden, que será el utilizado en este trabajo. Estos métodos se construyen bajo la suposición de procesos oscilatorios subyacentes y suponen, una alternativa a la estimación clásica del orden temporal de los instantes de muerte de los individuos a partir del ToD a partir marcadores clínicos (Li et al., 2013).

### 2.2.1 Método ORI (Inferencia con Restricciones de Orden).

El método ORI está basado en Inferencia con Restricciones de Orden. Esta metodología permite estimar el orden temporal buscando el orden circular óptimo entre los instantes de tiempo del conjunto de datos, entendido éste último como la permutación de los instantes de tiempo que en términos globales (para una medida de error) otorga mayor ritmicidad al conjunto total de genes con el que se esté trabajando.

Como ya se ha dicho, $\boldsymbol{X_j} = (X_{1j}, \dots, X_{nj})'$ representa la expresión correspondiente al gen $j, j = 1, \dots, n$; Denotaremos además como $D = \{\boldsymbol{X_j}\}_{j=1}^{J}$ la matriz de datos de expresión.

Por otro lado, $\Pi$ simboliza al conjunto de órdenes circulares posibles. Nótese que, para cada orden circular, $\boldsymbol{o} \in \Pi$; existe un modelo de señal circular de la forma $\boldsymbol{\mu} \in C_{\boldsymbol{o}}$ y que el número de órdenes circulares posibles es $(n-1)!$. La distancia entre un orden circular $\boldsymbol{o}$ y el conjunto de datos se define de la siguiente manera:

$$d(\boldsymbol{o},D) = \sum_{j=1}^{J} \sum_{i=1}^{n} v_j (X_{ij} - X^*_{o,ij})^2 \qquad (5)$$

Dónde $\boldsymbol{X}^*_{\boldsymbol{o}}$ es el estimador de Regresión Isotónica, (ver sección 2.3.1) bajo el orden circular $\boldsymbol{o}$. En concreto $\boldsymbol{X}^*_{\boldsymbol{o}}$, estima la señal up-down-up más próxima a los datos del gen $\boldsymbol{X_j}$ siguiendo el orden circular $\boldsymbol{o}$, mientras $v_j$ es el peso asociado al $j$-ésimo elemento.

Finalmente, el orden temporal se obtiene como solución del siguiente problema de optimización:

$$argmin_{o \in \Pi} d(\boldsymbol{o}, D) \qquad (6)$$

El orden óptimo será aquel que hace que globalmente todos los genes sean lo más cercanos posible a señales up-down-up. Se trata de un problema NP-hard ya que existen $\Pi = (n-1)!$ órdenes posibles y se obtiene una solución aproximada reformulando el Problema del Viajante (TSP) del siguiente modo:

1- Cada gen se representa como un grafo donde los nodos simbolizan los instantes de tiempo que se han de ordenar, mientras las aristas representan las distancias entre cada par de nodos.
2- A continuación, se calcula un grafo agregado, resultado de la suma ponderada de las distancias entre las aristas de cada grafo.
3- El problema queda reducido a recorrer todos los nodos de este grafo una única vez empleando la menor distancia posible, comenzando y terminando en un mismo nodo.
4- Este problema se resuelve de forma aproximada como un problema del viajante.
5- Finalmente, el orden óptimo obtenido junto con la formulación equivalente de señal circular en ambos espacios permite recuperar la ordenación de los instantes de tiempo que mejor recupera el carácter rítmico en la matriz de datos de expresión.

La Figura 2 contiene un esquema de la descripción del algoritmo que se acaba de detallar.

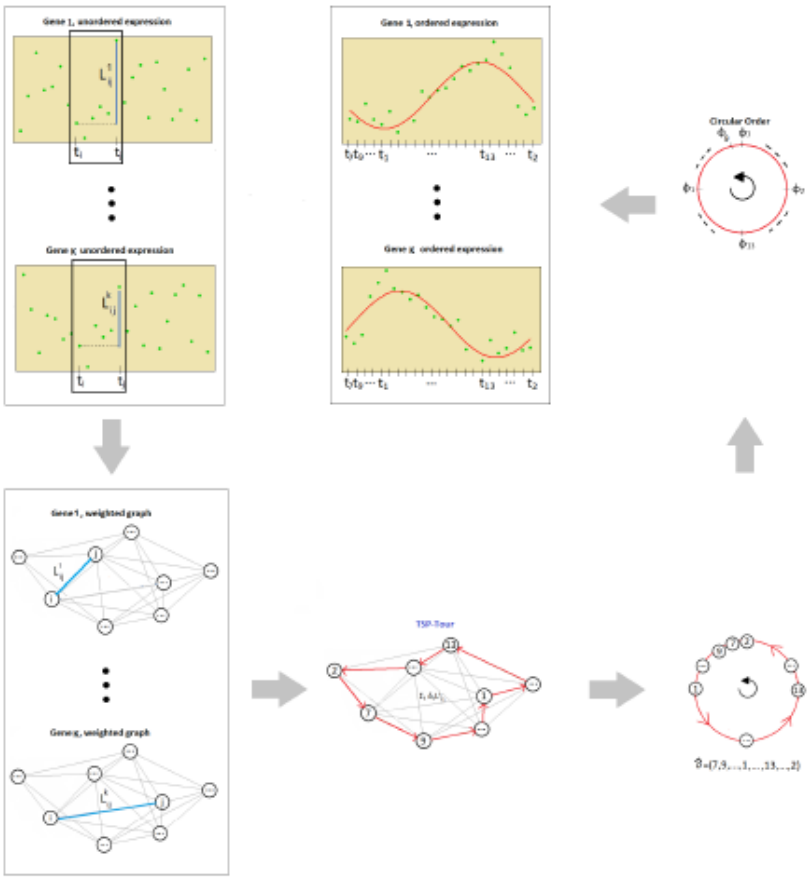


Figure S8: Temporal order estimation flowchart.

Figura 2: Imagen que representa el algoritmo empleado para estimar el orden mediante la metodología ORI. Fuente: Larriba, Rueda, Fernández, & Peddada, 2019.

## 2.3 Modelos de señal oscilatoria

Una vez obtenido el orden, se considera para su validación una colección de modelos para señales oscilatorias que están entre los más empleados en la literatura.

De forma general se definen todos los modelos como: $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, donde la señal up-down-up se especifica de forma diferente según el modelo empleado.

Nótese que $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ donde $\mu_i$ denota el valor de la señal en el instante $t_i$. Se asume además que $t_i \, \epsilon \, [0, 2\pi], i = 1, \dots, n$ . En otro caso se aplica la siguiente transformación, véase Rueda et al. (2019):

$$t' \epsilon \, [t_0, T + \, t_0], t = \frac{(t' - \, t_0) * 2\pi}{T} \tag{7}$$

Donde T es el periodo de los datos, por ejemplo 24 horas.

### 2.3.1 Modelo no paramétrico de Regresión Isotónica (IR)

Este modelo propuesto en Larriba et al. (2020) define la señal up-down-up ($\boldsymbol{\mu}$) como solución del siguiente problema de minimización:

$$\arg min_{z \epsilon c} \sum_{i=1}^{n} (X_i - Z_i)^2 \tag{8}$$

Nótese que la solución a este problema es $\boldsymbol{X}^*$, el estimador de Regresión Isotónica (IR) con respecto a $C$. Se trata del vector más próximo a los datos, en términos del error cuadrático, que verifica las restricciones de orden dadas en $C$, véase referencias Robertson et al. (1988) para más detalles.

Para derivar el estimador IR definido en (8), se ha diseñado un algoritmo computacionalmente eficiente basado en regresión isotónica y en los resultados teóricos descritos en Larriba et al. (2020). Los pasos de este algoritmo son:

1. En primer lugar, se debe encontrar $L^*$ y $U^*$que se buscarán entre los óptimos locales.
2. Se utilizará el algoritmo PAVA (Robertson et al., 1988) para calcular los estimadores de regresión isotónica estrictamente creciente (decrecientes) para cada una de las combinaciones posibles de $L^*$ y $U^*$.

3. Finalmente, se escoge aquella combinación de $L^*$ y $U^*$ que en el punto anterior tenga el menor error cuadrático medio.

La Figura 3 contiene una representación de la señal up-down-up considerada en este modelo.

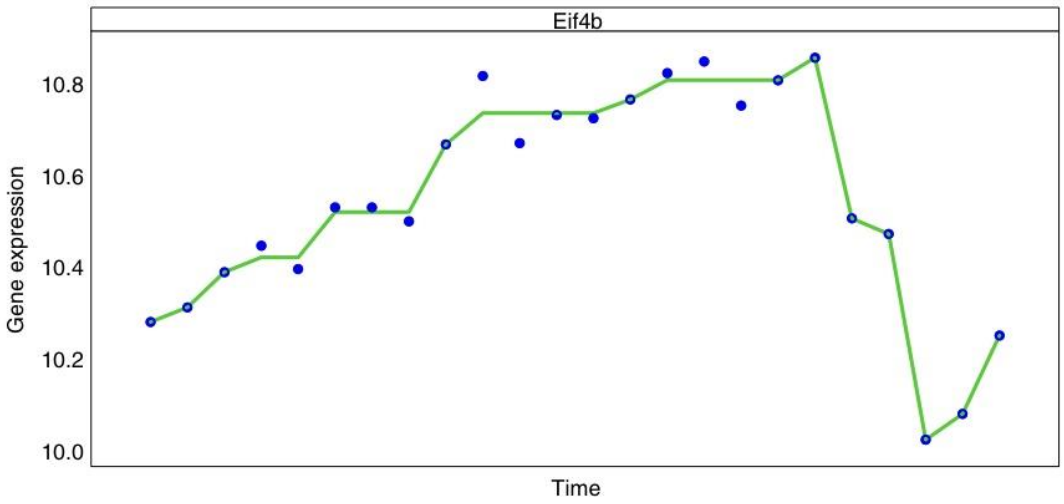


Figura 3: Modelo IR. Fuente Larriba et al. 2019, 2014.

### 2.3.2 Modelo paramétrico – Cosinor

Este modelo propuesto por Cornelissen (2014), define la señal up-down-up como la curva sinusoidal que mejor se ajusta a los datos en términos de mínimos cuadrados. La expresión paramétrica de la señal up-down-up propuesta en el modelo Cosinor es:

$$\mu(t) = M + Acos[t + \varphi],\ t \in [0, 2\pi] \quad (9)$$

Descripción de los parámetros empleados:

- *M*, también conocido como MESOR (*Midlin Estimating Statistic of Rhythm*), es el valor medio de la curva.
- *A* representa la amplitud de la curva, definida como la mitad de la distancia vertical entre su valor máximo y mínimo.
- Finalmente, $\varphi$ es la fase o acrofase. Representa el tiempo trascurrido desde el momento de referencia y el punto más elevado del ciclo.

La Figura 4 contiene una representación de la señal considerada en este modelo y de los parámetros del mismo.

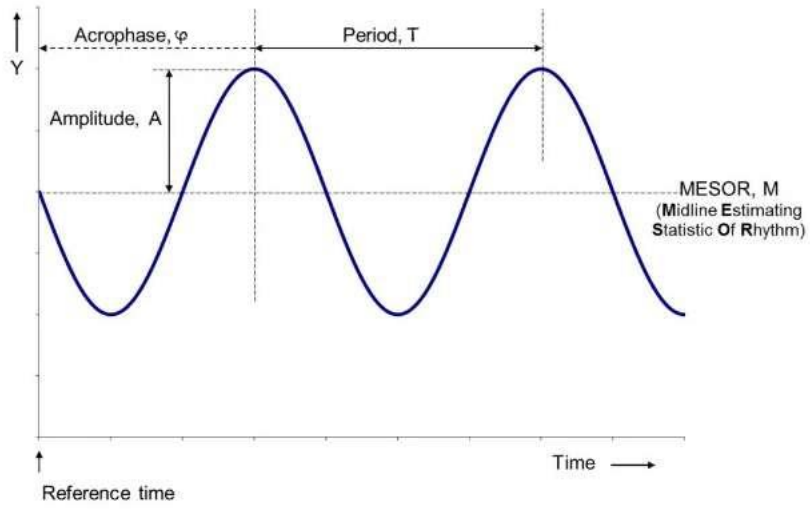


Figura 4: Modelo Cosinor y sus parámetros. Fuente Cornelissen, 2014.

### 2.3.3 Modelo paramétrico – FMM (*Frecuency Modulated Möbius*)

Este modelo de señal oscilatoria fue propuesto por Rueda et al. (2019). Se trata de un modelo que se adapta a una gran variedad de patrones al emplear una transformación de Möbius como función de enlace, en contraste con la función de enlace lineal que caracteriza al modelo Cosinor.
La señal up-down-up en el modelo FMM se define como:

$$\mu(t) = M + A\cos\big(\phi(t)\big) = M + Acos(\beta + 2\arctan(\omega tan\frac{t-\alpha}{2})), t\,\epsilon\,[0, 2\pi] \quad (10)$$

Descripción de los parámetros empleados:

- M es el *intercept* del modelo, $M \epsilon R$.
- A es la amplitud de la onda, $A \epsilon R^{+}$.
- $\alpha$ es el parámetro de localización de fase, $\alpha \in [0,2\pi]$ .
- $\beta$ es el parámetro de forma que define la asimetría. La onda es simétrica si $\beta$ = 0 y $\beta = \pi$, mientras que si toma un valor intermedio es asimétrica, $\beta \in [0,2\pi]$.
- $\omega$ es un parámetro de forma que define el apuntamiento de la onda. Si ω = 0, la onda es completamente apuntada, mientras que si ω = 1 estaríamos ante una sinusoidal, $\omega\,\epsilon\,[0,1]$.

En la siguiente figura ilustra una galería de señales up-down derivadas del modelo FMM, para distintos valores del par ($\omega$, $\beta$).

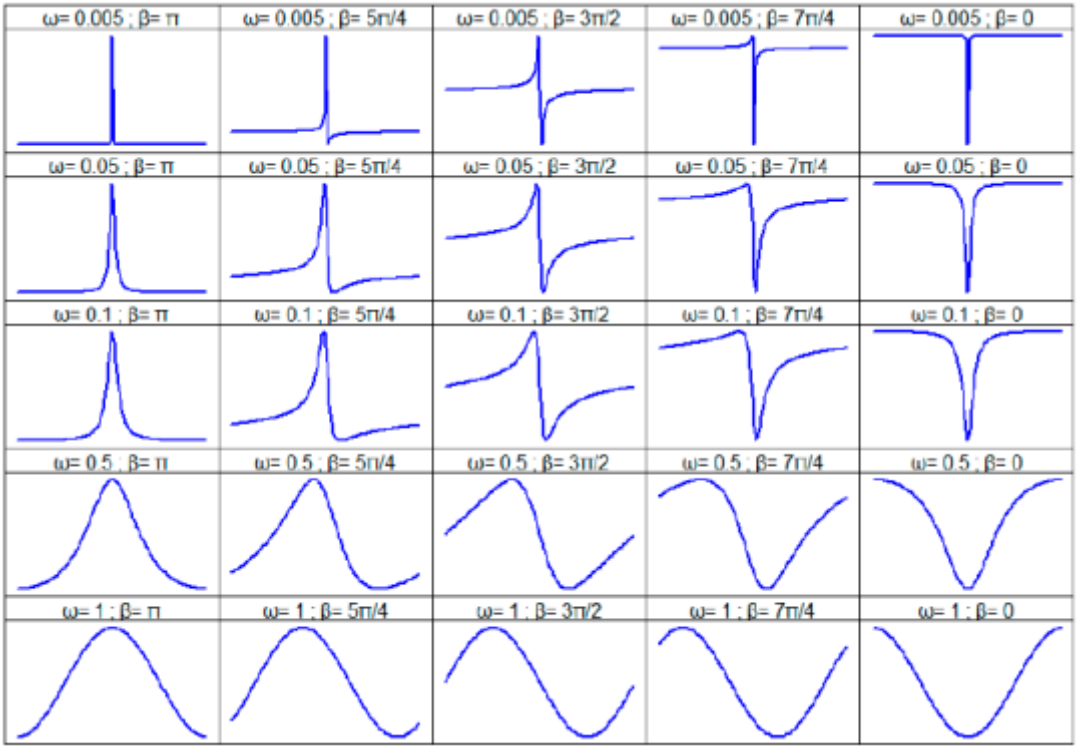


Figura 5: Influencia de los parámetros en el modelo FMM. Fuente: (Rueda, Larriba, & Peddada, 2019).

Otros parámetros de importancia en el modelo son los instantes de tiempo donde el modelo alcanza el máximo y el mínimo, $t_U$ y $t_L$, respectivamente, calculados como:

$$t_U = \alpha + 2\arctan\left(\frac{1}{\omega}\tan\left(\frac{-\beta}{2}\right)\right) \tag{11}$$

$$t_L = \alpha + 2\arctan\left(\frac{1}{\omega}\tan\left(\frac{\pi-\beta}{2}\right)\right) \tag{12}$$

## 2.4 Medidas de calidad de los modelos

Finalmente, en esta sección se describe la medida de bondad de ajuste empleada para la comparación de los modelos ajustados.

### 2.4.1 Coeficiente de determinación

Para medir la calidad del ajuste se ha empleado un coeficiente de determinación, $R^2$, similar a los utilizados en regresión lineal que refleja la bondad del ajuste de un modelo a la variable que se pretende explicar. El coeficiente de determinación oscila entre 0 y 1, cuanto más cercano a 1 se sitúe su valor, mayor será el ajuste del modelo.

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(\boldsymbol{X}-\widehat{\boldsymbol{X}})^2}{\sum_{j=1}^{n}(\boldsymbol{X}-\bar{X})^2} \qquad (13)$$

Donde $\widehat{\boldsymbol{X}}$ es el valor ajustado por el modelo para cada expresión del gen, $\boldsymbol{X}$ la expresión del gen y $\bar{X}$ el valor medio.

## 2.5. Datos direccionales

Cualquier fenómeno periódico es susceptible de ser representado en un soporte circular. En el caso estudiado, los ritmos circadianos o el orden obtenido mediante el método ORI, pueden visualizarse como direcciones en una circunferencia centrada en el origen.

A estos fines, se elige una dirección inicial y una orientación del círculo unidad. Con esta configuración, cada punto $x$ sobre la circunferencia tiene su correspondiente ángulo $\theta$.

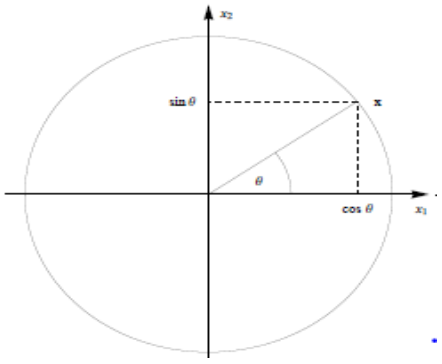


Figura 6: Representación de direcciones en coordenadas cartesianas y polares. Fuente: Gaussianos.com

De esta forma, partiendo de una posición inicial $\theta_0$, el punto $x$ describirá una rotación completa de $2p\pi$ radianes, con $p \epsilon N$, regresando a la misma posición. Este suceso puede expresarse como: $\theta = mod(\theta + 2p\pi)$.

Tal y como se anticipaba en el epígrafe introductorio de esta sección, los datos direccionales no admiten métodos de análisis estadístico lineal. Por ejemplo, la media habitual no tiene sentido puesto que la media de los valores circulares $10$ y $2\pi - 10$ no puede ser $\pi$ sino que debe ser $0$. Existe abundante literatura sobre este tema. Pueden consultarse entre otros los textos de Mardia & Jupp (2000) o de Fisher & Lee (1995). En este proyecto en concreto nos interesan los coeficientes de correlación entre datos direccionales. Se describen a continuación los dos más habituales que serán considerados en el análisis del problema que se está considerando.

### 2.5.1 Correlación de Jammaladamaka.

Sean $(\Phi, \Theta)$ variables aleatorias direccionales que representan ángulos con idéntica dirección de partida y sentido de rotación en un intervalo [0, 2π], con $\overline{\Phi}$ y $\overline{\Theta}$ como las direcciones medias de las distribuciones marginales de Φ y Θ respectivamente. El coeficiente de correlación de Jammalamadaka se define como:

$$\rho_c(\Phi, \Theta) = \frac{\mathrm{E}\sin(\Phi - \overline{\Phi})\sin(\Theta - \overline{\Theta})}{[\mathrm{E}\, sin^2(\Phi - \overline{\Phi})\ \mathrm{E}\ sin^2(\Theta - \overline{\Theta})]^{\frac{1}{2}}} \qquad (14)$$

Cabe remarcar que $\mathrm{E}\sin(\Phi - \overline{\Phi}) = \mathrm{E}\sin(\Theta - \overline{\Theta}) = 0$, (Jammalamadaka & Sarma, 1988).

Además, este coeficiente satisface las siguientes propiedades:

1. $\rho_c(\Phi, \Theta)$ no depende de la primera dirección empleada para cada variable.
2. $\rho_c(\Phi, \Theta) = \rho_c(\Theta, \Phi)$.
3. $-1 \leq \rho_c(\Phi, \Theta) \leq 1$.
4. $\rho_c(\Phi, \Theta) = 0$ si Φ y Θ son independientes, aunque el recíproco no es cierto.
5. Si Φ y Θ están totalmente correlacionados, entonces $\rho_c(\Phi, \Theta) = 1$ si y solo si $\Phi = \Theta + const(mod\ 2\pi)$ y $\rho_c(\Phi, \Theta) = -1$ si y solo si $\Phi = -\Theta + const(mod\ 2\pi)$.

Para una muestra $(\phi_i, \theta_i),\ i = 1, \dots, n$ de una distribución circular definida en el intervalo [0, 2π]; el coeficiente de correlación de Jammalamadaka muestral se expresa como:

$$r_{c,n} = \frac{\sum_{1 \le i < j \le n} \sin(\phi_i - \bar{\phi}_n)\sin(\theta_i - \bar{\theta}_n)}{[\sum_{1 \le i < j \le n} sin^2\,(\phi_i - \bar{\phi}_n) \sum_{1 \le i < j \le n} sin^2\,(\theta_i - \bar{\theta}_n)]^{\frac{1}{2}}} \tag{15}$$

Donde $\bar{\phi}_n$ y $\bar{\theta}_n$ son las medias muestrales circulares de la muestra aleatoria simple $(\phi_i, \theta_i),\ i = 1, \dots, n$ respectivamente.

Teorema 1: (Jammalamadaka & Sarma, 1988)

Si ninguna de las distribuciones marginales es uniforme, entonces

$\sqrt{n}(r_{c,n} - \rho_c(\Phi, \Theta))$ converge a una distribución $\mathrm{N}(0, \sigma^2)$ donde:

$$\begin{aligned} \sigma^2 = {} & \frac{\lambda_{22}}{\lambda_{20}\lambda_{02}} - \rho_c(\Phi, \Theta)\left[\frac{\lambda_{13}}{\lambda_{20}\sqrt{\lambda_{20}\lambda_{02}}} + \frac{\lambda_{31}}{\lambda_{02}\sqrt{\lambda_{20}\lambda_{02}}}\right] \\ & + \frac{\rho_c^2(\Phi,\Theta)}{4}\left[1 + \frac{\lambda_{40}}{\lambda_{20}^2} + \frac{\lambda_{04}}{\lambda_{02}^2} + \frac{\lambda_{22}}{\lambda_{20}\lambda_{02}}\right] \end{aligned} \tag{16}$$

Remarcar que si $n$ es lo suficientemente grande, se puede estimar $\lambda_{ij}$ como:

$$\widehat{\lambda_{ij}} = \frac{1}{n}\sum_{l=1}^{n} sin^i(\phi_l - \bar{\phi}_n) sin^j(\theta_l - \bar{\theta}_n), i, j = 1, \dots, n. \tag{17}$$

Para más detalle, véase Jammalamadaka & Sarma (1998).

Intervalo de confianza:

Es posible estimar un intervalo de confianza para $\rho_c(\Phi, \Theta)$ gracias al Teorema 1. Por estandarización, se obtiene la variable aleatoria:

$$Z = \frac{\sqrt{n}\left(r_{c,n} - \rho_c(\Phi,\Theta)\right)}{\sqrt{\sigma^2}} \sim \mathrm{N}(0, 1) \tag{18}$$

Por lo tanto, es posible hallar los valores $-z$ y $z$, entre los cuales se encuentra $Z$, con probabilidad $1 - \alpha$. Para una confianza del 95%, se obtiene:

$$P(-z \leq Z \leq z) = 1 - \alpha = 0.95 \tag{19}$$

Donde valor se obtiene como:

$$\begin{gathered} \mathrm{F}(z) = P(Z \leq z) = 1 - \frac{\alpha}{2} = 0.975; \\ z = \mathrm{F}^{-1}(0.975) = 1.96 \end{gathered} \tag{20}$$

Si se sustituye en (19):

$$P\left(-1.96 < \frac{\sqrt{n}\left(r_{c,n} - \rho_c(\Phi, \Theta)\right)}{\sqrt{\sigma^2}} < 1.96\right) = 1 - \alpha \tag{21}$$

Se obtiene el siguiente intervalo de confianza para $\rho_c(\Phi, \Theta)$ con una confianza del 95%:

$$P\left(r_{c,n} - \frac{1.96 * \sqrt{\sigma^2}}{\sqrt{n}} < \rho_c(\Phi, \Theta) < r_{c,n} + \frac{1.96 * \sqrt{\sigma^2}}{\sqrt{n}}\right) = 1 - \alpha \tag{22}$$

Donde $\sigma^2$ se podrá estimar por $\widehat{\sigma^2}$ através de la estimación dada para $\lambda_{ij}$ y sustituyendo $\rho_c(\Phi, \Theta)$ por $r_{c,n}$.

### 2.5.2 Asociación T-Lineal: Fisher & Lee

Sean $(\Phi, \Theta)$ variables aleatorias direccionales, un modelo para representar la asociación lineal entre variables reales es el siguiente:

$$\Phi = \Theta + \theta_0 \, (mod\ 2\pi) \text{ o } \Phi = -\Theta + \theta_0 \, (mod\ 2\pi) \qquad (23)$$

Este modelo es conocido como asociación T-Lineal, que describe asociación positiva y negativa entre dos variables circulares.

Para una muestra $(\theta_i, \phi_i),\ i = 1, \dots, n$ de una distribución circular definida en el intervalo $[0, 2\pi]$, Fisher y Lee en (Fisher & Lee, 1995), proponen utilizar el siguiente coeficiente de correlación muestral para valorar la posible existencia de esta asociación:

$$\widehat{\rho_T} = \frac{\sum_{1 \le i < j \le n} \sin(\theta_i - \theta_j)\sin(\phi_i - \phi_j)}{[\sum_{1 \le i < j \le n} sin^2\,(\theta_i - \theta_j) \sum_{1 \le i < j \le n} sin^2\,(\phi_i - \phi_j)]^{\frac{1}{2}}} \qquad (24)$$

Cuanto más cercano se encuentre $\widehat{\rho_T}$ de 1 o -1, mayor será el grado de asociación. La hipótesis de no asociación T-Lineal se rechaza si $\widehat{\rho_T}$ difiere mucho de 0.

# 3. DATOS

El conjunto de datos empleados para la realización de este proyecto proviene de una muestra postmortem de 104 pacientes anónimos, de los cuales se extrajeron secuencias de ADN de la corteza cerebral. Este procedimiento fue realizado por la Universidad de Pittsburgh y la Escuela de medicina Monte Sinaí en base a los siguientes criterios descritos en Seney et al. (2019):

- Sujetos con una hora de muerte conocida (TOD) y con un lapso de dos horas entre el evento desencadenante de la muerte y el fallecimiento.
- Sujetos con una edad menor de 65 años.
- Sujetos con un intervalo postmortem (tiempo transcurrido desde la muerte de una persona) menor de 30 horas.
- Sujetos que no hayan padecido ningún tipo de enfermedad psicológica.

Por otra parte, este proyecto se centra en las siguientes variables dicotómicas:

- Sexo: hombre y mujer.
- Origen de los datos: MSSM y Pitt.
- Causa de la muerte: causa 1 (muerte cardiovascular) y causa 5 (otros); las causas 2 y 4 no se tendrán en cuenta en el análisis ya que solo hay un individuo que ha fallecido por cáncer no cerebral y otro por EPOC.

La Tabla 1 muestra la frecuencia absoluta y relativa de las variables mencionadas previamente.

| | **SEXO** | | **INSTITUCION** | | **CAUSA DE LA MUERTE** | | | |
|---|---|---|---|---|---|---|---|---|
| | HOMBRE | MUJER | MSSM | PITT | 1 | 2 | 4 | 5 |
| ABSOLUTA | 81 | 23 | 43 | 61 | 70 | 1 | 1 | 32 |
| RELATIVA | 78% | 22% | 41% | 59% | 67% | 1% | 1% | 31% |

Tabla 1: Distribución de los individuos por sexo, origen de los datos y causa de la muerte

| **EQUIVALENCIA - CAUSA DE LA MUERTE** | |
|---|---|
| **1** | **CARDIO VASCULAR** |
| **2** | **CÁNCER NO CEREBRAL** |
| **4** | **ENFERMEDAD PULMONAR OBTRUCTIVA CRÓNICA - EPOC** |
| **5** | **OTROS** |

Tabla 2: Definición de la causa de la muerte

Los siguientes diagramas de barras representan la distribución de los individuos según el sexo, origen de los datos y la causa de la muerte.

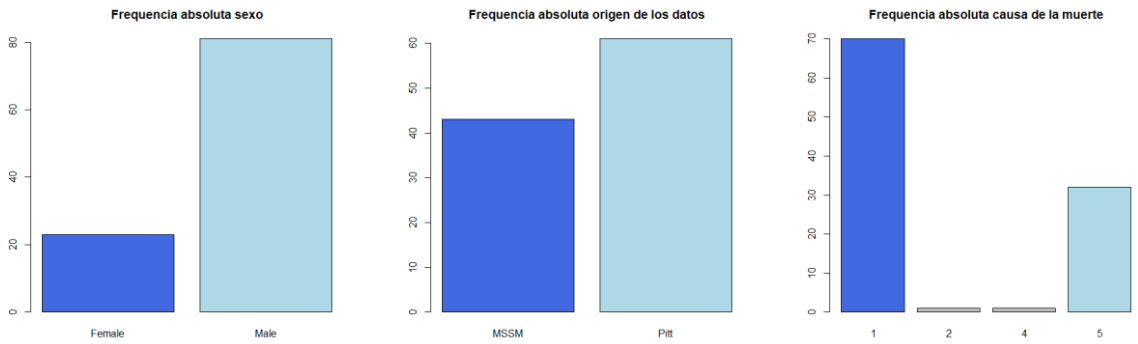


Figura 7: diagramas de barras para la distribución del sexo, origen de los datos y edad

Asimismo, para completar el análisis descriptivo de los datos, se muestran los histogramas de las variables continuas edad y la hora de la muerte.

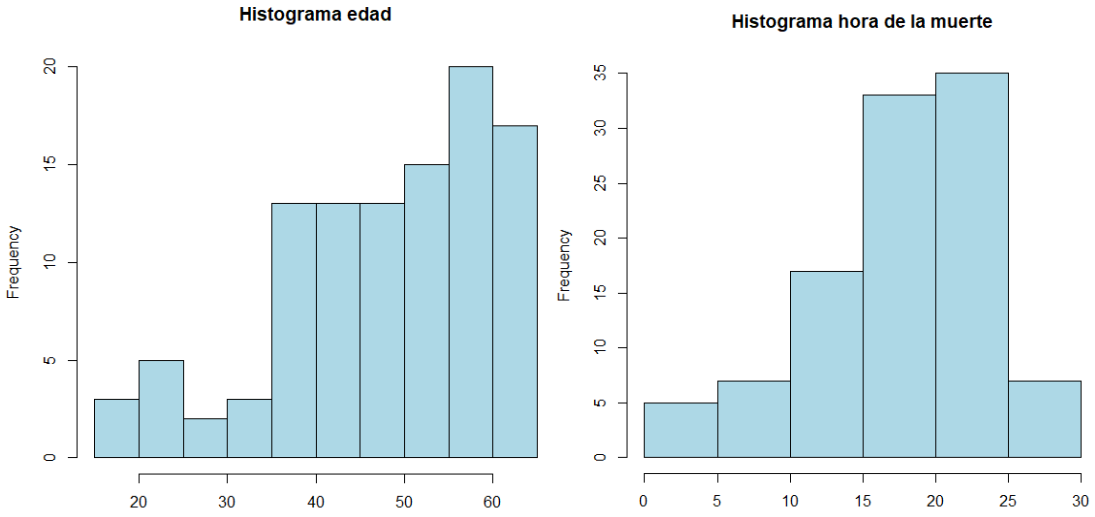


Figura 8: histogramas de la edad y hora de la muerte

Del total de individuos se ha recogido la expresión de 13914 genes, seleccionando una submuestra de 25 genes *core*, empleados con asiduidad en el campo de la cronobiología y cuya ritmicidad ya ha sido valorada en otros estudios como Mure et al. (2018), concretamente:

| GENES | | | | |
|---|---|---|---|---|
| ARNTL | NR1D1 | PER1 | PER2 | ACAP3 |
| NPAS2 | BHLHE41 | PER3 | MXD4 | CPTP |
| CLOCK | NR1D2 | TEF | NELFA | SKI |
| NFIL3 | DBP | HLF | FGFRL1 | FAM213B |
| CRY1 | CIART | CRY2 | IDUA | PRDM16 |

Tabla 3: submuestra de 25 genes core

## 3.1 Tratamiento de los datos

Con la finalidad de homogeneizar los datos para comparar simultáneamente modelos y órdenes, es preciso realizar un reescalado por cada eje del plano cartesiano.

La transformación relativa al eje de ordenadas consiste en escalar los datos al intervalo [-1,1]. Para ello, se utiliza la siguiente expresión:

$$datoEscalado = \frac{x - xmin}{xmax - \ xmin} \tag{25}$$

A continuación, se lleva a cabo un segundo reescalado en el eje de abscisas, ya que la muestra de datos original se encuentra en un intervalo [-6, 18] asociado a la escala ambiental Zeitgeber, usada habitualmente para sincronizar datos horarios de distintos husos. Mientras que los modelos de ajuste implementados en este proyecto, tales como el Cosinor, el FMM y el no paramétrico, requieren que los tiempos ToD estén comprendidos en el intervalo [0, 2π].

Con esta finalidad, se aplica la siguiente expresión:

$$datoEscalado = \ 2\pi \frac{x - xmin}{xmax - xmin} \tag{26}$$

## 3.2 Sincronización de los datos

Para facilitar la visualización de los genes, se realiza una traslación de los datos. Concretamente, se toma como referencia el gen ARNTL ajustado al modelo no paramétrico para cada variable categórica, trasladándolo tantas posiciones como sean necesarias para que el *peak*, momento de máxima expresión del gen, del modelo quede representado en el centro del gráfico. Finalmente, esta traslación de aplica al resto de genes.

Cabe destacar que se toma como referencia el gen ARNTL puesto que en los mamíferos este gen juega un papel vital en la regulación del metabolismo y la homeostasis de la glucosa, teniendo su momento de máxima expresión antes de la puesta de sol. (Ruben et al., 2018)

## 3.3 Terminología

En este apartado, se especifica la equivalencia terminológica entre los conceptos empleados a lo largo del documento y sus correspondientes acrónimos o expresiones.

- Peak: momento de máxima expresión del gen.
- NP: modelo no paramétrico de Regresión Isotónica.
- FMM: *Frecuency Modulated Möbius.*
- Variable dicotómica: variable que solo puede tomar dos valores.
- MSSM y PITT: instituciones Monte Sinaí y Pittsburgh, respectivamente.
- ToD: Time of Death.

# 4. ANALISIS Y RESULTADOS

En esta sección se presentan los resultados obtenidos tras la aplicación de las técnicas mencionadas en la metodología, así como su correspondiente análisis y las conclusiones alcanzadas.

En primer lugar, se muestran los genes con mejor ajuste para el conjunto de los individuos y para cada variable dicotómica, así como la comparación entre modelos.

A continuación, se aborda el problema fundamental que se trata en este trabajo. Se estudia la posible influencia de estas variables dicotómicas en la configuración de los *peaks,* es decir si la ordenación de los *peaks* en el ciclo circadiano cambia con las categorías de las variables o no.

Conviene recordar que, como ya se ha dicho en el apartado anterior, se están considerando tres variables categóricas: el sexo, la causa de muerte y el origen de los datos. Parece más que razonable la hipótesis de que el origen de los datos no influye en la ordenación de los *peaks* por lo que se utilizará esta variable dicotómica como banco de pruebas de la metodología empleada. Además, se tomarán los resultados del ajuste Cosinor como modelo de referencia puesto que es el modelo más ampliamente utilizado en la literatura.

## 4.1 Generación de los órdenes y sincronización.

En primer lugar, se generan los órdenes con el método ORI, de acuerdo a lo desarrollado en la sección 2.2.1 del capítulo 2 de metodología. Existen métodos alternativos como la estimación del orden ToD, frecuentemente empleada en la literatura, que ordenan los individuos de manera ascendente según su hora estimada de muerte. Sin embargo, esta última metodología resulta, en términos de error, globalmente inferior al orden obtenido por el método ORI, cuyo error cuadrático medio es menor, como puede verse en la Figura 9. Además, las medidas de bondad de ajuste presentan globalmente valores más reducidos para el orden ToD, véase Prieto (2021), aunque en este trabajo no se tenía en cuenta el efecto de las variables categóricas, como sí se hace aquí.

| Control MSE | | | |
|---|---|---|---|
| | ORI | ORI Reducido | ToD |
| NP | 0.0645 | 0.0827 | 0.1164 |
| FMM | 0.0937 | 0.1118 | 0.1343 |
| Cosinor | 0.1056 | 0.1217 | 0.1455 |

Figura 9: Imagen que muestra el error (MSE) del método ORI, ORI reducido, y ToD para los datos globales considerados en este estudio. Fuente: TFG de Carlota Prieto.

Con el método ORI se generan dos órdenes diferentes para cada variable dicotómica: sexo, causa de la muerte y origen de los datos; partiendo del conjunto de 25 genes *core* mencionados en la sección 3.2.

Destacar que el método ORI solo proporciona la ordenación óptima entre los individuos, pero para poder estimar los modelos paramétricos es necesario que estos se encuentren en un instante de tiempo exacto. En consecuencia, los individuos son equiespaciados en el intervalo $[0, 2\pi]$.

Una vez calculados todos los órdenes, se ajustan los correspondientes modelos no paramétricos y se centran los *peaks* del modelo tomando como referencia el gen ARNTL. Como consecuencia, se produce la traslación de las distintas observaciones de cada gen, así como para todos los órdenes. En Tabla 4 se muestran los *peaks* del modelo no paramétrico y en la Figura 10 se representan los genes ARNTL, NPAS2 y CLOCK antes y después de la sincronización, con el ajuste del modelo Cosinor en color rojo, el FMM en azul y el no paramétrico en verde. Esta leyenda de colores se aplicará en todos los gráficos posteriores.

| | **PEAK NP ANTES** | **PEAK NP DESPUÉS** |
|---|---|---|
| ARNTL | 1,268 | 3,141 |
| NPAS2 | 3,685 | 5,558 |
| CLOCK | 5,618 | 1,208 |

Tabla 4: Comparación de los Peaks del modelo NP antes y después de la sincronización

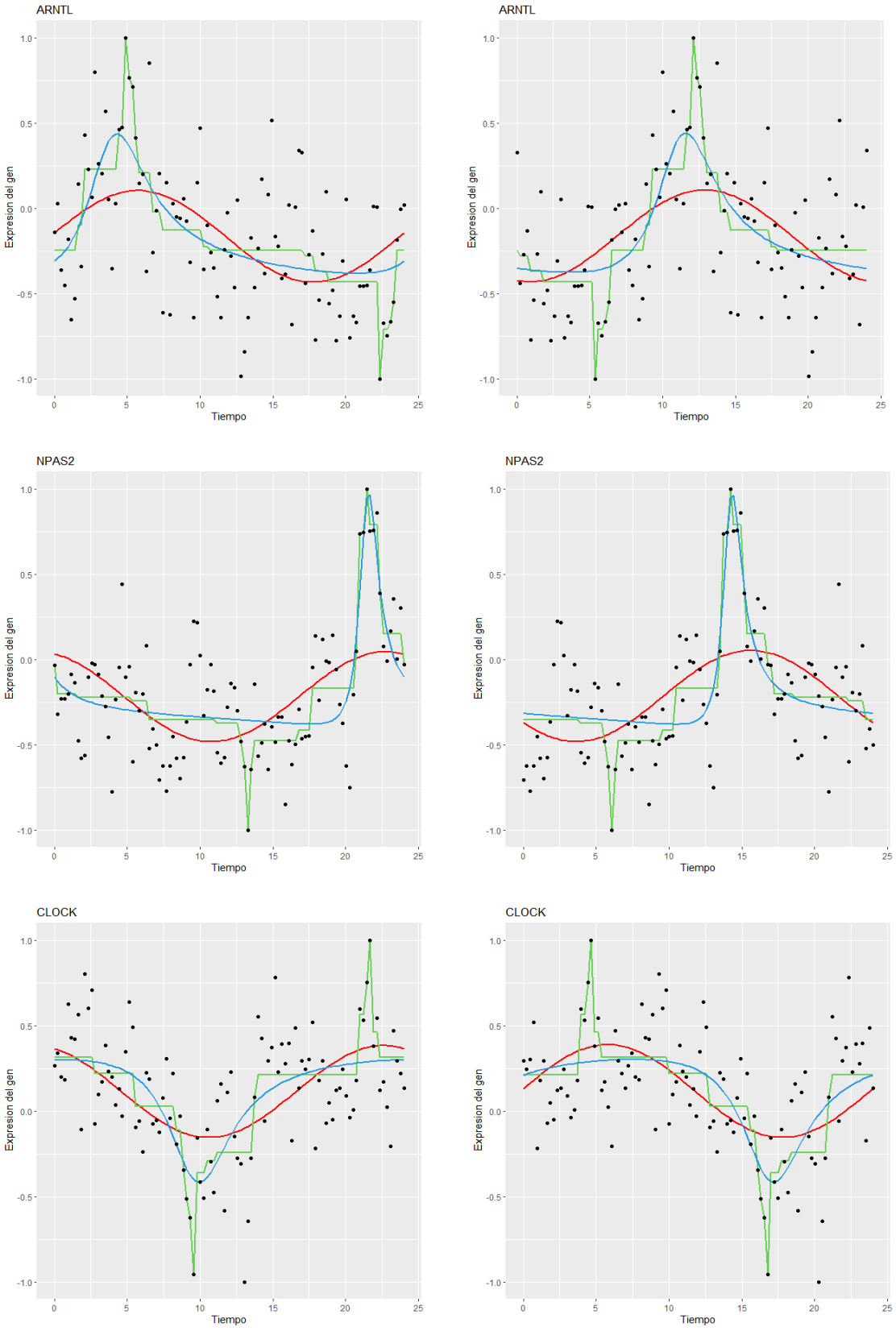


Figura 10: Genes ARNTL, NPAS2 y CLOCK antes y después de la sincronización.

## 4.2 Ajuste de los modelos y resultados

Tras obtener los órdenes y la traslación correspondientes, se ajustan los datos de expresión de cada gen mediante los modelos Cosinor y FMM. Además, se obtienen los respectivos coeficientes de determinación y *peaks* de cada modelo con el objetivo de estudiar los ritmos de cada gen y comparar los ajustes para cada variable dicotómica.

En la Tabla 5, se muestran los $R^2$ obtenidos a partir del orden generado para todos los individuos considerando los modelos sobre el conjunto total de los individuos:

| GEN | FMM | COS | NP |
|---|---|---|---|
| ARNTL | 0,344 | 0,208 | 0,521 |
| NPAS2 | 0,524 | 0,237 | 0,671 |
| CLOCK | 0,372 | 0,283 | 0,605 |
| NFIL3 | 0,242 | 0,038 | 0,501 |
| CRY1 | 0,239 | 0,214 | 0,491 |
| NR1D1 | 0,62 | 0,587 | 0,814 |
| BHLHE41 | 0,396 | 0,361 | 0,627 |
| NR1D2 | 0,326 | 0,321 | 0,645 |
| DBP | 0,472 | 0,461 | 0,702 |
| CIART | 0,251 | 0,221 | 0,532 |
| PER1 | 0,713 | 0,662 | 0,855 |
| PER3 | 0,519 | 0,497 | 0,724 |
| TEF | 0,325 | 0,317 | 0,552 |
| HLF | 0,216 | 0,176 | 0,436 |
| CRY2 | 0,429 | 0,429 | 0,65 |
| PER2 | 0,393 | 0,332 | 0,618 |
| MXD4 | 0,553 | 0,541 | 0,754 |
| NELFA | 0,611 | 0,582 | 0,746 |
| FGFRL1 | 0,472 | 0,458 | 0,689 |
| IDUA | 0,656 | 0,646 | 0,847 |
| ACAP3 | 0,574 | 0,564 | 0,748 |
| CPTP | 0,531 | 0,526 | 0,757 |
| SKI | 0,646 | 0,548 | 0,819 |
| FAM213B | 0,552 | 0,532 | 0,771 |
| PRDM16 | 0,346 | 0,299 | 0,641 |

Tabla 5: Valores de $R^2$para los ajustes de los genes core de todos los individuos para los diferentes modelos considerados.

En la Tabla 6, se muestran los *peaks* obtenidos a partir del orden generado para todos los individuos considerando los modelos sobre el conjunto total de los individuos:

| GEN | FMM | COS | NP |
|---|---|---|---|
| ARNTL | 3,032 | 3,375 | 3,142 |
| NPAS2 | 5,647 | 5,966 | 5,558 |
| CLOCK | 1,985 | 1,517 | 1,208 |
| NFIL3 | 0,076 | 1,908 | 0,060 |
| CRY1 | 2,722 | 2,327 | 2,960 |
| NR1D1 | 5,322 | 5,753 | 5,437 |
| BHLHE41 | 6,137 | 0,474 | 0,060 |
| NR1D2 | 5,508 | 0,775 | 1,148 |
| DBP | 5,409 | 5,373 | 5,377 |
| CIART | 0,226 | 6,172 | 5,256 |
| PER1 | 5,364 | 5,881 | 5,437 |
| PER3 | 5,946 | 0,038 | 5,498 |
| TEF | 5,359 | 5,763 | 0,121 |
| HLF | 6,257 | 1,094 | 1,208 |
| CRY2 | 5,423 | 5,493 | 5,377 |
| PER2 | 5,321 | 5,988 | 5,558 |
| MXD4 | 4,790 | 4,907 | 5,377 |
| NELFA | 5,230 | 5,123 | 5,437 |
| FGFRL1 | 4,689 | 4,974 | 4,652 |
| IDUA | 5,510 | 5,363 | 5,437 |
| ACAP3 | 5,283 | 5,057 | 5,377 |
| CPTP | 4,880 | 4,884 | 5,377 |
| SKI | 5,572 | 5,663 | 5,437 |
| FAM213B | 5,035 | 5,147 | 5,377 |
| PRDM16 | 4,801 | 5,465 | 5,317 |

Tabla 6: Valores de los *peaks* para los ajustes de los genes core de todos los individuos para los diferentes modelos considerados.

Los genes que presentan mejor ajuste para todos los modelos son: NR1D1, PER1, NELFA e IDUA. Estos genes tienen un $R^2$ superior a 0,6 exceptuando los genes NR1D1 y NELFA que para el modelo Cosinor tienen un $R^2$ de 0,582. La siguiente Tabla 7 muestra los $R^2$ y *peaks* de los genes mencionados junto a su representación gráfica.

| | NR1D1 | | PER1 | | NELFA | | IDUA | |
|---|---|---|---|---|---|---|---|---|
| | **R2** | **PEAKS** | **R2** | **PEAKS** | **R2** | **PEAKS** | **R2** | **PEAKS** |
| **FMM** | 0,62 | 5,753 | 0,713 | 5,88 | 0,611 | 5,122 | 0,656 | 5,362 |
| **COSINOR** | 0,587 | 5,322 | 0,662 | 5,36 | 0,582 | 5,229 | 0,646 | 5,51 |
| **NP** | 0,814 | 5,437 | 0,855 | 5,44 | 0,746 | 5,437 | 0,847 | 5,437 |

Tabla 7: Valores de $R^2$ y peaks para los ajustes de los genes NR1D1, PER1, NELFA E IDUA de todos los individuos para los diferentes modelos considerados.

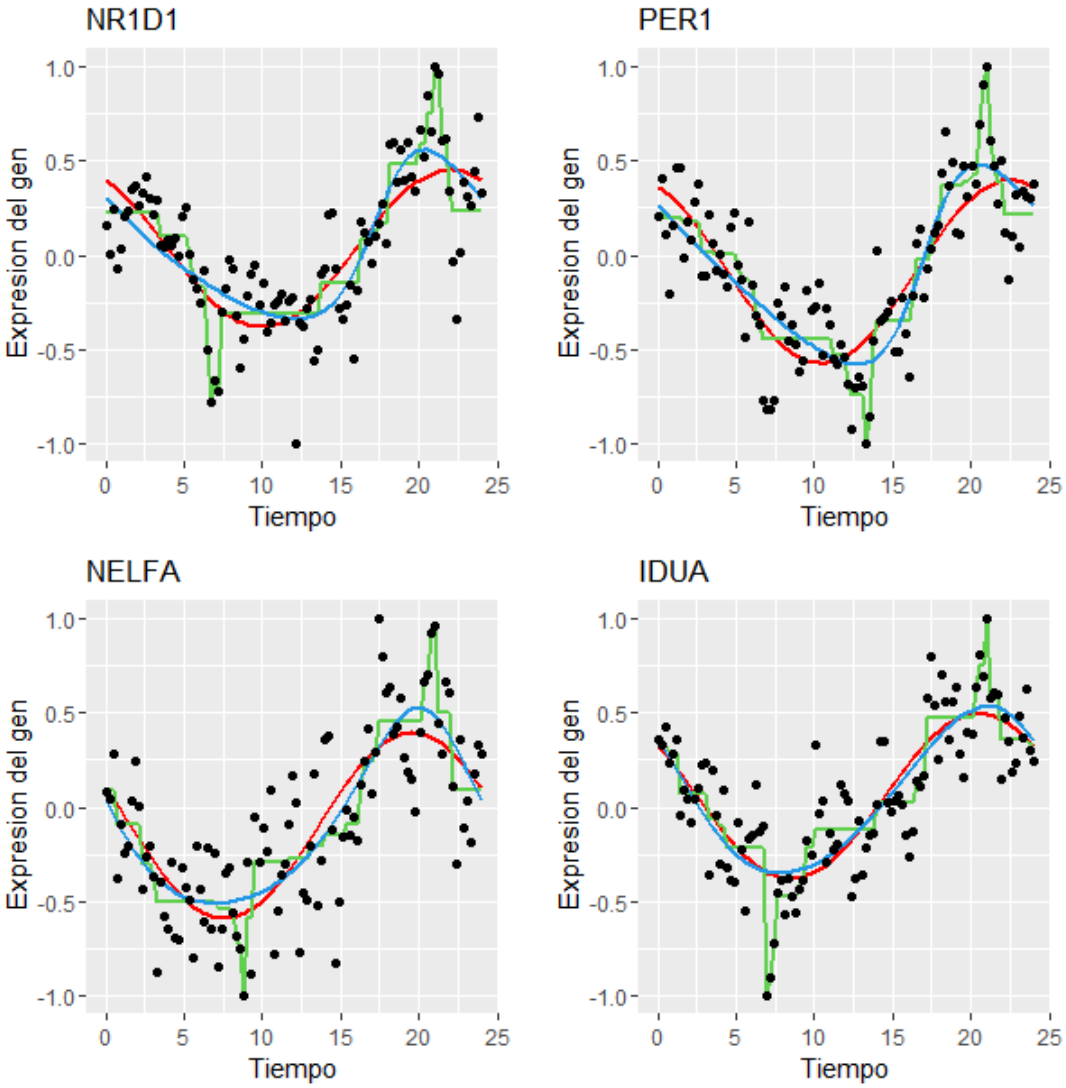


Figura 11: Genes NR1D1, PER1, NELFA e IDUA de todos los individuos con diferentes modelos ajustados.

A continuación, para cada variable dicotómica, se muestra el ajuste de todos los genes para los modelos considerados junto con la selección de genes con mayor $R^2$, con sus respectivos *peaks* y representación gráfica.

## Tabla $R^2$ sexo

| | HOMBRE | | | MUJER | | |
|---|---|---|---|---|---|---|
| **GEN** | **FMM** | **COS** | **NP** | **FMM** | **COS** | **NP** |
| ARNTL | 0,294 | 0,188 | 0,478 | 0,455 | 0,142 | 0,57 |
| NPAS2 | 0,427 | 0,172 | 0,608 | 0,665 | 0,221 | 0,826 |
| CLOCK | 0,459 | 0,367 | 0,694 | 0,551 | 0,424 | 0,824 |
| NFIL3 | 0,303 | 0,112 | 0,529 | 0,704 | 0,215 | 0,833 |
| CRY1 | 0,422 | 0,346 | 0,575 | 0,685 | 0,42 | 0,874 |
| NR1D1 | 0,792 | 0,725 | 0,927 | 0,725 | 0,708 | 0,949 |
| BHLHE41 | 0,387 | 0,356 | 0,636 | 0,734 | 0,594 | 0,939 |
| NR1D2 | 0,39 | 0,377 | 0,698 | 0,546 | 0,395 | 0,809 |
| DBP | 0,524 | 0,518 | 0,742 | 0,718 | 0,617 | 0,893 |
| CIART | 0,394 | 0,292 | 0,586 | 0,564 | 0,284 | 0,778 |
| PER1 | 0,751 | 0,697 | 0,899 | 0,723 | 0,686 | 0,887 |
| PER3 | 0,462 | 0,411 | 0,712 | 0,721 | 0,705 | 0,905 |
| TEF | 0,245 | 0,235 | 0,453 | 0,444 | 0,257 | 0,594 |
| HLF | 0,253 | 0,235 | 0,479 | 0,545 | 0,395 | 0,779 |
| CRY2 | 0,504 | 0,478 | 0,707 | 0,458 | 0,309 | 0,671 |
| PER2 | 0,485 | 0,353 | 0,66 | 0,336 | 0,14 | 0,614 |
| MXD4 | 0,623 | 0,623 | 0,825 | 0,82 | 0,788 | 0,936 |
| NELFA | 0,643 | 0,565 | 0,789 | 0,629 | 0,581 | 0,831 |
| FGFRL1 | 0,483 | 0,474 | 0,754 | 0,582 | 0,408 | 0,781 |
| IDUA | 0,634 | 0,582 | 0,83 | 0,714 | 0,645 | 0,903 |
| ACAP3 | 0,615 | 0,582 | 0,792 | 0,655 | 0,519 | 0,801 |
| CPTP | 0,545 | 0,518 | 0,794 | 0,679 | 0,575 | 0,84 |
| SKI | 0,657 | 0,518 | 0,815 | 0,724 | 0,551 | 0,893 |
| FAM213B | 0,619 | 0,546 | 0,824 | 0,523 | 0,296 | 0,682 |
| PRDM16 | 0,445 | 0,382 | 0,72 | 0,274 | 0,151 | 0,519 |

Tabla 8: Valores de $R^2$para los ajustes de los genes de la variable dicotómica sexo para los diferentes modelos considerados.

**Tabla *peaks* sexo**

| | HOMBRE | | | MUJER | | |
|---|---|---|---|---|---|---|
| **GEN** | **FMM** | **COS** | **NP** | **FMM** | **COS** | **NP** |
| ARNTL | 2,893 | 3,212 | 3,025 | 3,528 | 1,782 | 3,005 |
| NPAS2 | 5,169 | 5,958 | 5,197 | 5,790 | 5,156 | 5,464 |
| CLOCK | 2,238 | 1,660 | 1,396 | 1,186 | 0,133 | 1,639 |
| NFIL3 | 0,639 | 1,881 | 0,776 | 4,711 | 0,021 | 4,644 |
| CRY1 | 3,155 | 2,480 | 3,336 | 5,471 | 0,981 | 1,366 |
| NR1D1 | 5,248 | 5,794 | 5,042 | 4,599 | 4,495 | 4,644 |
| BHLHE41 | 1,005 | 0,541 | 0,776 | 4,750 | 5,610 | 4,644 |
| NR1D2 | 1,230 | 0,837 | 1,474 | 4,092 | 5,412 | 4,644 |
| DBP | 5,179 | 5,372 | 4,964 | 4,184 | 4,066 | 4,098 |
| CIART | 0,214 | 6,091 | 6,206 | 3,777 | 4,475 | 4,098 |
| PER1 | 5,350 | 5,870 | 5,042 | 5,073 | 4,668 | 4,098 |
| PER3 | 5,287 | 6,272 | 5,120 | 5,133 | 5,222 | 4,917 |
| TEF | 5,285 | 5,712 | 0,698 | 5,897 | 4,147 | 5,464 |
| HLF | 0,771 | 1,284 | 1,396 | 4,948 | 0,192 | 5,737 |
| CRY2 | 5,086 | 5,440 | 4,964 | 2,361 | 3,934 | 4,371 |
| PER2 | 5,237 | 5,918 | 5,197 | 5,767 | 5,227 | 5,464 |
| MXD4 | 4,920 | 4,920 | 4,964 | 3,318 | 3,684 | 3,551 |
| NELFA | 5,091 | 5,221 | 5,042 | 3,570 | 4,141 | 2,732 |
| FGFRL1 | 4,860 | 5,038 | 4,887 | 2,819 | 3,579 | 2,732 |
| IDUA | 5,188 | 5,366 | 5,042 | 4,587 | 4,238 | 3,005 |
| ACAP3 | 5,280 | 5,014 | 4,964 | 3,220 | 4,040 | 3,278 |
| CPTP | 4,997 | 4,955 | 4,964 | 2,856 | 3,531 | 3,005 |
| SKI | 5,358 | 5,652 | 5,042 | 5,458 | 4,653 | 5,464 |
| FAM213B | 4,950 | 5,170 | 4,732 | 2,544 | 3,873 | 2,732 |
| PRDM16 | 0,339 | 5,483 | 4,887 | 6,271 | 3,879 | 4,098 |

Tabla 9: Valores de los *peaks* para los ajustes de los genes de la variable dicotómica sexo para los diferentes modelos considerados.

La Tabla 8 y la Tabla 9 contienen los resultados relativos al sexo. Los $R^2$ son similares a los del conjunto total de individuos, presentando un mejor ajuste para todos los modelos los genes: NR1D1, PER1, MXD4 e IDUA. Estos genes tienen un $R^2$ superior a 0,6 exceptuando el gen IDUA que para el modelo Cosinor presenta un $R^2$ de 0,582.

## Gen NR1D1

| | HOMBRE | | MUJER | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,792 | 5,248 | 0,725 | 4,599 |
| **COSINOR** | 0,725 | 5,794 | 0,708 | 4,495 |
| **NP** | 0,927 | 5,042 | 0,949 | 4,644 |

Tabla 10: Valores de $R^2$ y peaks para los ajustes del gen NR1D1 de la variable dicotómica sexo para los diferentes modelos considerados.

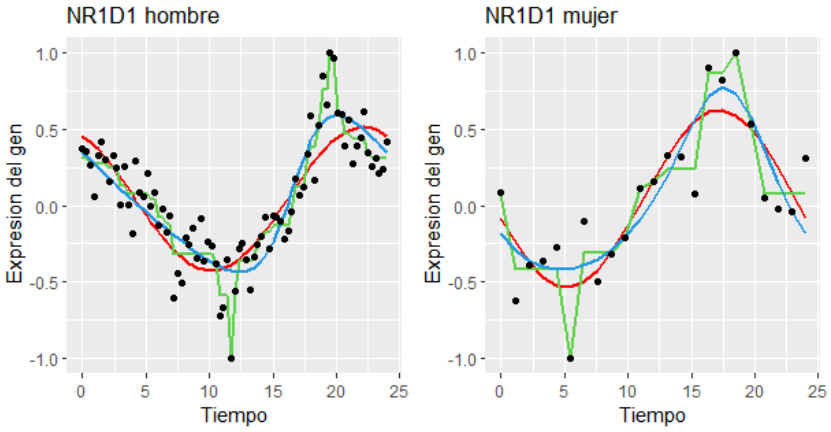


Figura 12: Comparación del gen NR1D1 para la variable dicotómica sexo.

## Gen PER1

| | HOMBRE | | MUJER | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,751 | 5,35 | 0,723 | 5,073 |
| **COSINOR** | 0,697 | 5,87 | 0,686 | 4,668 |
| **NP** | 0,899 | 5,042 | 0,887 | 4,098 |

Tabla 11: Valores de $R^2$ y peaks para los ajustes del gen PER1 de la variable dicotómica sexo para los diferentes modelos considerados.

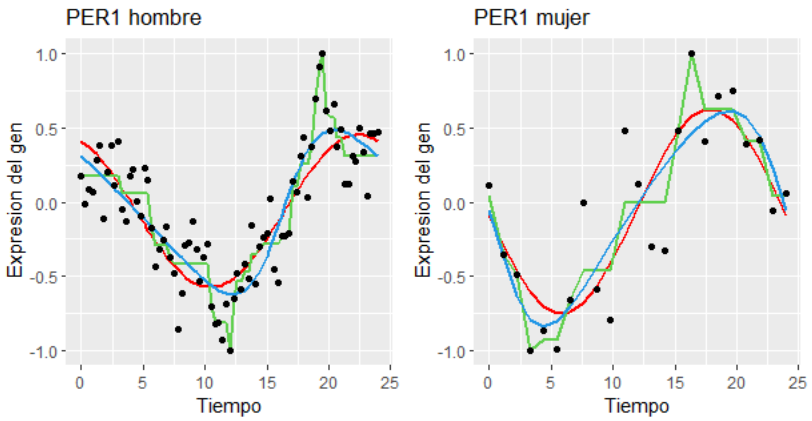


Figura 13: Comparación del gen PER1 para la variable dicotómica sexo.

## Gen MXD4

| | HOMBRE | | MUJER | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,623 | 4,92 | 0,82 | 3,318 |
| **COSINOR** | 0,623 | 4,92 | 0,788 | 3,684 |
| **NP** | 0,825 | 4,964 | 0,936 | 3,551 |

Tabla 12: Valores de $R^2$ y peaks para los ajustes del gen MXD4 de la variable dicotómica sexo para los diferentes modelos considerados.

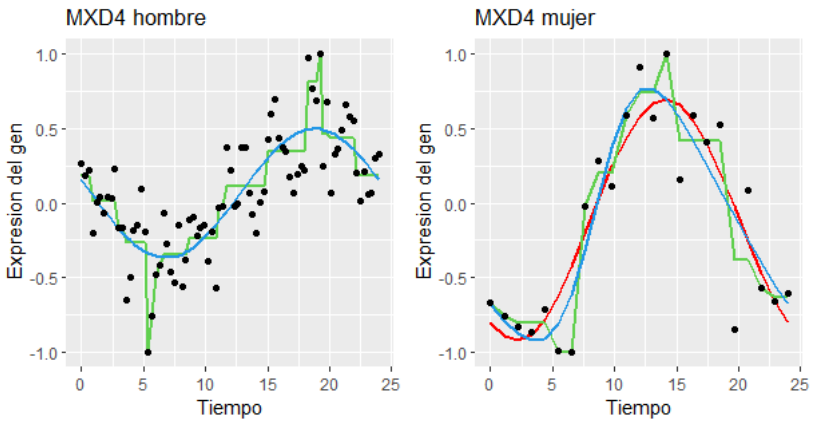


Figura 14: Comparación del gen MXD4 para la variable dicotómica sexo.

Señalar que en el gráfico de los hombres del gen MXD4, el ajuste del modelo FMM está superpuesto sobre el modelo Cosinor.

## Gen IDUA

| | HOMBRE | | MUJER | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,634 | 5,188 | 0,714 | 4,587 |
| **COSINOR** | 0,582 | 5,366 | 0,645 | 4,238 |
| **NP** | 0,83 | 5,042 | 0,903 | 3,005 |

Tabla 13: Valores de $R^2$ y peaks para los ajustes del gen IDUA de la variable dicotómica sexo para los diferentes modelos considerados.

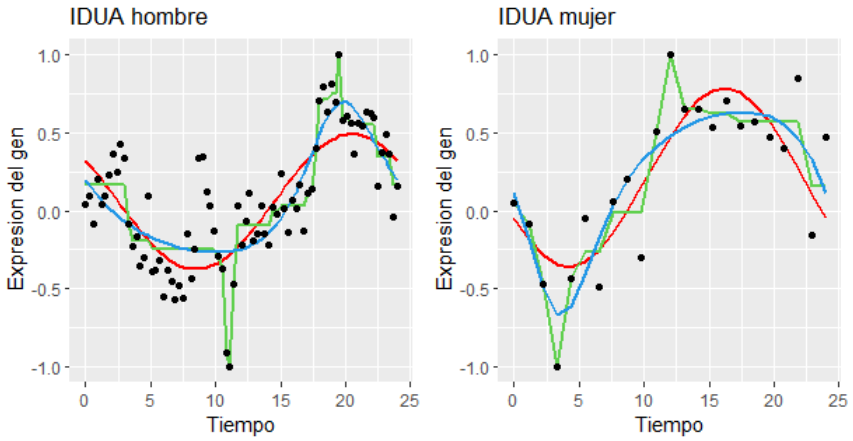


Figura 15: Comparación del gen IDUA para la variable dicotómica sexo.

Gráficamente, puede observarse que los patrones de expresión up-down-up entre ambos sexos son muy similares, distinguiéndose únicamente por una amplitud de la onda superior en el caso de las mujeres, especialmente en el gen MXD4, que podría deberse a que se cuenta con un menor número de observaciones del género femenino.

## Tabla $R^2$origen de los datos

| | MSSM | | | PITT | | |
|---|---|---|---|---|---|---|
| **GEN** | **FMM** | **COS** | **NP** | **FMM** | **COS** | **NP** |
| ARNTL | 0,359 | 0,071 | 0,552 | 0,339 | 0,233 | 0,58 |
| NPAS2 | 0,592 | 0,254 | 0,795 | 0,561 | 0,319 | 0,775 |
| CLOCK | 0,312 | 0,136 | 0,545 | 0,609 | 0,544 | 0,778 |
| NFIL3 | 0,295 | 0,168 | 0,585 | 0,375 | 0,159 | 0,626 |
| CRY1 | 0,26 | 0,253 | 0,528 | 0,48 | 0,45 | 0,69 |
| NR1D1 | 0,813 | 0,751 | 0,956 | 0,731 | 0,712 | 0,888 |
| BHLHE41 | 0,454 | 0,437 | 0,787 | 0,476 | 0,469 | 0,68 |
| NR1D2 | 0,305 | 0,09 | 0,514 | 0,434 | 0,391 | 0,676 |
| DBP | 0,518 | 0,474 | 0,737 | 0,6 | 0,545 | 0,819 |
| CIART | 0,5 | 0,298 | 0,704 | 0,445 | 0,252 | 0,65 |
| PER1 | 0,78 | 0,738 | 0,944 | 0,759 | 0,754 | 0,9 |
| PER3 | 0,353 | 0,32 | 0,702 | 0,668 | 0,656 | 0,859 |
| TEF | 0,413 | 0,382 | 0,669 | 0,35 | 0,314 | 0,598 |
| HLF | 0,329 | 0,108 | 0,52 | 0,407 | 0,333 | 0,644 |
| CRY2 | 0,425 | 0,414 | 0,639 | 0,509 | 0,503 | 0,803 |
| PER2 | 0,383 | 0,287 | 0,597 | 0,536 | 0,508 | 0,747 |
| MXD4 | 0,564 | 0,443 | 0,758 | 0,605 | 0,605 | 0,845 |
| NELFA | 0,695 | 0,54 | 0,841 | 0,55 | 0,518 | 0,783 |
| FGFRL1 | 0,541 | 0,362 | 0,835 | 0,559 | 0,515 | 0,722 |
| IDUA | 0,75 | 0,596 | 0,904 | 0,497 | 0,488 | 0,796 |
| ACAP3 | 0,605 | 0,397 | 0,826 | 0,519 | 0,519 | 0,788 |
| CPTP | 0,619 | 0,354 | 0,766 | 0,597 | 0,593 | 0,823 |
| SKI | 0,796 | 0,57 | 0,934 | 0,575 | 0,5 | 0,799 |
| FAM213B | 0,667 | 0,433 | 0,854 | 0,611 | 0,56 | 0,86 |
| PRDM16 | 0,395 | 0,347 | 0,694 | 0,354 | 0,314 | 0,671 |

Tabla 14: Valores de $R^2$para los ajustes de los genes de la variable dicotómica institución para los diferentes modelos considerados.

## Tabla *peaks* origen de los datos

| | MSSM | | | PITT | | |
|---|---|---|---|---|---|---|
| **GEN** | **FMM** | **COS** | **NP** | **FMM** | **COS** | **NP** |
| ARNTL | 3,072 | 3,390 | 3,069 | 2,349 | 2,960 | 3,090 |
| NPAS2 | 0,039 | 0,164 | 5,991 | 5,646 | 6,101 | 5,768 |
| CLOCK | 5,208 | 3,777 | 2,630 | 5,379 | 4,786 | 5,150 |
| NFIL3 | 3,624 | 4,734 | 4,968 | 3,610 | 4,222 | 5,562 |
| CRY1 | 3,074 | 3,193 | 3,361 | 4,486 | 4,063 | 4,738 |
| NR1D1 | 6,165 | 6,259 | 6,137 | 0,731 | 0,461 | 0,927 |
| BHLHE41 | 5,098 | 5,413 | 4,968 | 5,769 | 5,626 | 5,150 |
| NR1D2 | 4,646 | 5,463 | 4,530 | 6,104 | 5,393 | 4,429 |
| DBP | 1,141 | 0,359 | 0,731 | 0,752 | 1,002 | 1,030 |
| CIART | 1,274 | 6,225 | 0,731 | 0,756 | 0,346 | 0,824 |
| PER1 | 5,821 | 6,226 | 6,137 | 0,450 | 0,291 | 1,030 |
| PER3 | 5,975 | 0,033 | 0,000 | 5,651 | 5,964 | 5,768 |
| TEF | 5,559 | 5,999 | 5,114 | 1,124 | 0,855 | 1,133 |
| HLF | 5,443 | 4,662 | 5,260 | 5,729 | 5,054 | 5,665 |
| CRY2 | 5,684 | 0,311 | 0,877 | 0,745 | 0,731 | 1,030 |
| PER2 | 0,140 | 0,703 | 0,292 | 5,794 | 6,225 | 5,768 |
| MXD4 | 5,913 | 0,481 | 0,000 | 1,430 | 1,430 | 1,030 |
| NELFA | 5,971 | 0,055 | 6,137 | 1,146 | 1,253 | 1,133 |
| FGFRL1 | 5,634 | 6,106 | 5,406 | 1,317 | 1,507 | 1,442 |
| IDUA | 5,774 | 0,267 | 6,137 | 0,963 | 0,851 | 0,927 |
| ACAP3 | 5,983 | 0,506 | 6,137 | 1,208 | 1,208 | 1,030 |
| CPTP | 5,683 | 0,170 | 5,553 | 1,322 | 1,406 | 1,030 |
| SKI | 6,240 | 0,003 | 6,137 | 6,038 | 0,412 | 5,768 |
| FAM213B | 5,767 | 0,321 | 5,553 | 1,151 | 1,050 | 1,133 |
| PRDM16 | 2,125 | 0,270 | 5,845 | 1,096 | 0,847 | 0,927 |

Tabla 15: Valores de los *peaks* para los ajustes de los genes de la variable dicotómica institución para los diferentes modelos considerados.

La Tabla 14 y la Tabla 15 muestran los resultados relativos al origen de los datos. Los genes que presentan mejor ajuste para todos los modelos son: NR1D1, PER1, NELFA e SKI. Estos genes tienen un $R^2$ superior a 0,55.

### Gen NR1D1

| | MSSM | | PITT | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,813 | 6,165 | 0,731 | 0,731 |
| **COSINOR** | 0,751 | 6,259 | 0,712 | 0,461 |
| **NP** | 0,956 | 6,137 | 0,888 | 0,927 |

Tabla 16: Valores de $R^2$ y peaks para los ajustes del gen NR1D1 de la variable dicotómica institución para los diferentes modelos.

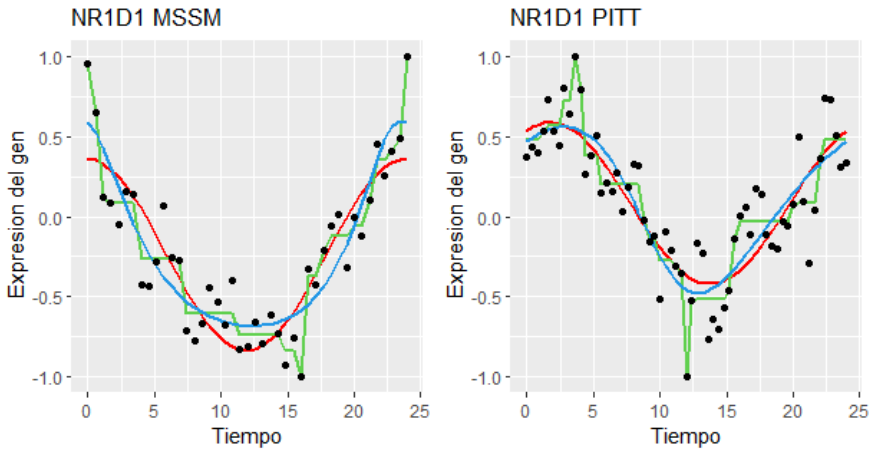


Figura 16: Comparación del gen NR1D1 para la variable dicotómica institución.

## Gen PER1

| | MSSM | | PITT | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,78 | 5,821 | 0,759 | 0,45 |
| **COSINOR** | 0,738 | 6,226 | 0,754 | 0,291 |
| **NP** | 0,944 | 6,137 | 0,9 | 1,03 |

Tabla 17: Valores de $R^2$ y peaks para los ajustes del gen PER1 de la variable dicotómica institución para los diferentes modelos considerados.

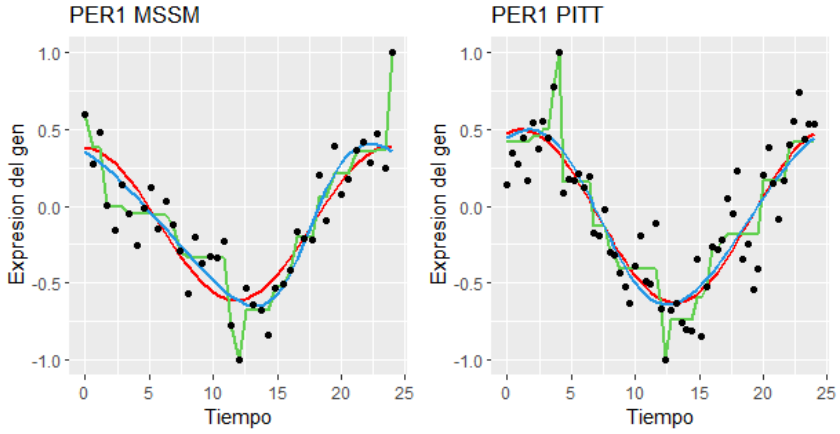


Figura 17: Comparación del gen PER1 para la variable dicotómica institución.

## Gen NELFA

| | MSSM | | PITT | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,695 | 5,971 | 0,55 | 1,146 |
| **COSINOR** | 0,54 | 0,055 | 0,518 | 1,253 |
| **NP** | 0,841 | 6,137 | 0,783 | 1,133 |

Tabla 18: Valores de $R^2$ y peaks para los ajustes del gen NELFA de la variable dicotómica institución para los diferentes modelos considerados.

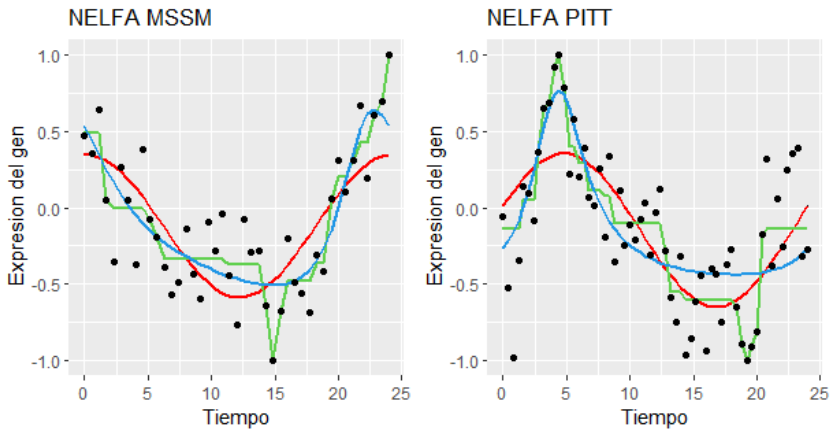


Figura 18: Comparación del gen NELFA para la variable dicotómica institución.

## Gen SKI

| | MSSM | | PITT | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,796 | 6,24 | 0,575 | 6,038 |
| **COSINOR** | 0,57 | 0,003 | 0,5 | 0,412 |
| **NP** | 0,934 | 6,137 | 0,799 | 5,768 |

Tabla 19: Valores de $R^2$ y peaks para los ajustes del gen SKI de la variable dicotómica institución para los diferentes modelos considerados.

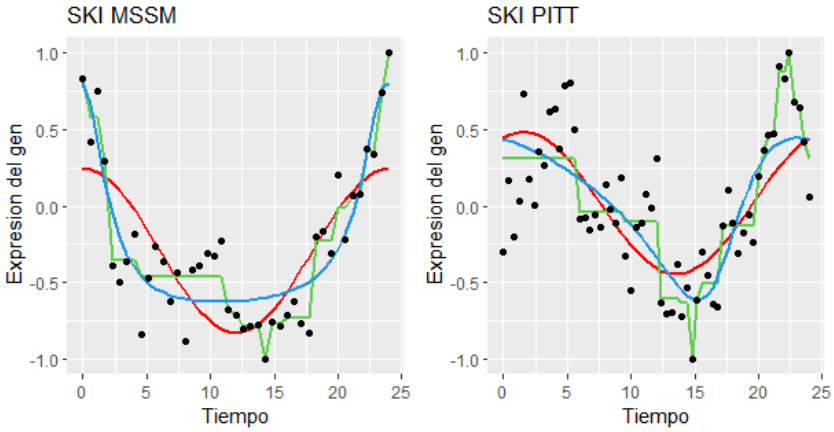


Figura 19: Comparación del gen SKI para la variable dicotómica institución.

Al igual que sucede con el sexo, el origen de los datos presenta patrones muy similares. Además, los *peaks* se encuentran en idénticos instantes temporales.

### Tabla $R^2$ causa de la muerte

| | CAUSA 1 | | | CAUSA 5 | | |
|---|---|---|---|---|---|---|
| **GEN** | **FMM** | **COS** | **NP** | **FMM** | **COS** | **NP** |
| ARNTL | 0,234 | 0,121 | 0,441 | 0,403 | 0,133 | 0,577 |
| NPAS2 | 0,677 | 0,534 | 0,823 | 0,598 | 0,16 | 0,78 |
| CLOCK | 0,376 | 0,338 | 0,678 | 0,584 | 0,292 | 0,829 |
| NFIL3 | 0,367 | 0,27 | 0,592 | 0,344 | 0,123 | 0,775 |
| CRY1 | 0,421 | 0,367 | 0,673 | 0,498 | 0,426 | 0,752 |
| NR1D1 | 0,678 | 0,678 | 0,877 | 0,747 | 0,63 | 0,937 |
| BHLHE41 | 0,357 | 0,332 | 0,544 | 0,432 | 0,411 | 0,855 |
| NR1D2 | 0,459 | 0,268 | 0,698 | 0,648 | 0,629 | 0,795 |
| DBP | 0,662 | 0,574 | 0,825 | 0,556 | 0,498 | 0,833 |
| CIART | 0,553 | 0,384 | 0,76 | 0,53 | 0,439 | 0,711 |
| PER1 | 0,731 | 0,703 | 0,887 | 0,651 | 0,624 | 0,848 |
| PER3 | 0,434 | 0,412 | 0,655 | 0,609 | 0,423 | 0,843 |
| TEF | 0,228 | 0,219 | 0,47 | 0,452 | 0,405 | 0,66 |
| HLF | 0,315 | 0,273 | 0,593 | 0,248 | 0,099 | 0,406 |
| CRY2 | 0,432 | 0,393 | 0,691 | 0,615 | 0,546 | 0,8 |
| PER2 | 0,44 | 0,342 | 0,639 | 0,505 | 0,305 | 0,774 |
| MXD4 | 0,533 | 0,317 | 0,777 | 0,777 | 0,671 | 0,93 |
| NELFA | 0,437 | 0,237 | 0,644 | 0,77 | 0,677 | 0,897 |
| FGFRL1 | 0,457 | 0,156 | 0,669 | 0,544 | 0,417 | 0,737 |
| IDUA | 0,64 | 0,485 | 0,859 | 0,804 | 0,781 | 0,957 |
| ACAP3 | 0,555 | 0,283 | 0,739 | 0,678 | 0,662 | 0,849 |
| CPTP | 0,479 | 0,285 | 0,759 | 0,624 | 0,572 | 0,901 |
| SKI | 0,77 | 0,623 | 0,887 | 0,675 | 0,64 | 0,91 |
| FAM213B | 0,493 | 0,317 | 0,718 | 0,752 | 0,676 | 0,927 |
| PRDM16 | 0,314 | 0,237 | 0,568 | 0,426 | 0,353 | 0,796 |

Tabla 20: Valores de $R^2$para los ajustes de los genes de la variable dicotómica causa de la muerte para los diferentes modelos considerados.

**Tabla *peaks* causa de la muerte**

| | CAUSA 1 | | | CAUSA 5 | | |
|---|---|---|---|---|---|---|
| **GEN** | **FMM** | **COS** | **NP** | **FMM** | **COS** | **NP** |
| ARNTL | 3,149 | 3,147 | 3,142 | 0,234 | 0,121 | 0,441 |
| NPAS2 | 1,244 | 1,020 | 1,346 | 0,677 | 0,534 | 0,823 |
| CLOCK | 1,966 | 1,992 | 1,975 | 0,376 | 0,338 | 0,678 |
| NFIL3 | 1,387 | 2,249 | 0,898 | 0,367 | 0,270 | 0,592 |
| CRY1 | 3,887 | 3,107 | 3,590 | 0,421 | 0,367 | 0,673 |
| NR1D1 | 0,280 | 0,280 | 0,808 | 0,678 | 0,678 | 0,877 |
| BHLHE41 | 1,173 | 1,201 | 0,808 | 0,357 | 0,332 | 0,544 |
| NR1D2 | 5,370 | 1,160 | 2,064 | 0,459 | 0,268 | 0,698 |
| DBP | 0,491 | 5,921 | 0,718 | 0,662 | 0,574 | 0,825 |
| CIART | 5,671 | 6,242 | 5,745 | 0,553 | 0,384 | 0,760 |
| PER1 | 0,725 | 0,397 | 0,718 | 0,731 | 0,703 | 0,887 |
| PER3 | 1,359 | 0,756 | 0,808 | 0,434 | 0,412 | 0,655 |
| TEF | 0,627 | 0,377 | 0,718 | 0,228 | 0,219 | 0,470 |
| HLF | 1,095 | 1,641 | 1,975 | 0,315 | 0,273 | 0,593 |
| CRY2 | 0,506 | 6,239 | 0,718 | 0,432 | 0,393 | 0,691 |
| PER2 | 1,514 | 0,634 | 1,167 | 0,440 | 0,342 | 0,639 |
| MXD4 | 0,586 | 5,761 | 0,718 | 0,533 | 0,317 | 0,777 |
| NELFA | 0,892 | 0,077 | 0,718 | 0,437 | 0,237 | 0,644 |
| FGFRL1 | 0,944 | 5,582 | 0,359 | 0,457 | 0,156 | 0,669 |
| IDUA | 0,948 | 0,199 | 0,269 | 0,640 | 0,485 | 0,859 |
| ACAP3 | 1,265 | 6,223 | 0,718 | 0,555 | 0,283 | 0,739 |
| CPTP | 0,552 | 5,740 | 0,718 | 0,479 | 0,285 | 0,759 |
| SKI | 0,992 | 0,697 | 0,808 | 0,770 | 0,623 | 0,887 |
| FAM213B | 0,940 | 6,053 | 0,718 | 0,493 | 0,317 | 0,718 |
| PRDM16 | 1,159 | 5,698 | 0,359 | 0,314 | 0,237 | 0,568 |

Tabla 21: Valores de los *peaks* para los ajustes de los genes de la variable dicotómica causa de la muerte para los diferentes modelos considerados.

La Tabla 20 y la Tabla 21 contienen los resultados relativos a la causa de muerte. Los genes que presentan mejor ajuste son: NR1D1, PER1, IDUA, y SKI, con un $R^2$ superior a 0,6 exceptuando el gen IDUA que para el modelo Cosinor tienen un $R^2$ de 0,48.

**Gen NR1D1**

| | CAUSA 1 | | CAUSA 5 | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,678 | 0,28 | 0,747 | 4,582 |
| **COSINOR** | 0,678 | 0,28 | 0,63 | 5,028 |
| **NP** | 0,877 | 0,808 | 0,937 | 4,516 |

Tabla 22: Valores de $R^2$ y peaks para los ajustes del gen NR1D1 de la variable dicotómica causa de la muerte para los modelos.

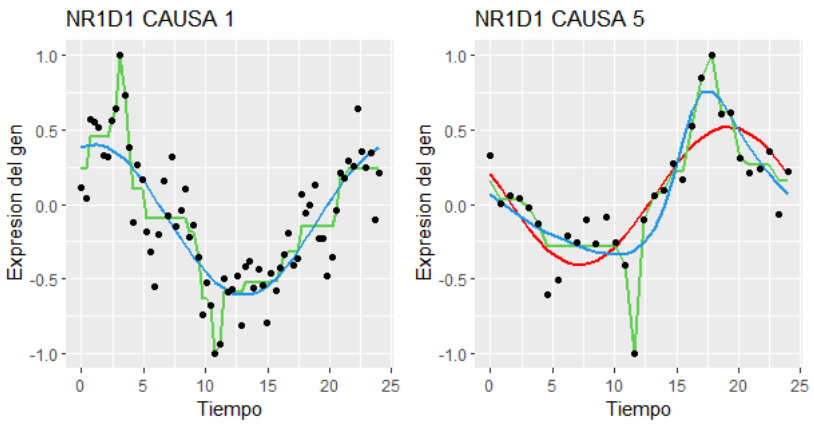


Figura 20: Comparación del gen NR1D1 para la variable dicotómica causa de la muerte.

## Gen PER1

| | CAUSA 1 | | CAUSA 5 | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,731 | 0,725 | 0,651 | 4,685 |
| **COSINOR** | 0,703 | 0,397 | 0,624 | 4,973 |
| **NP** | 0,887 | 0,718 | 0,848 | 4,516 |

Tabla 23: Valores de $R^2$ y peaks para los ajustes del gen PER1 de la variable dicotómica causa de la muerte para los diferentes modelos considerados.

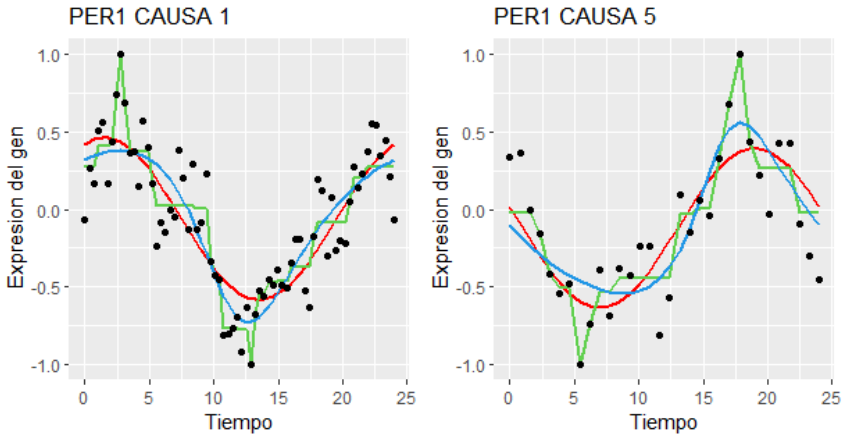


Figura 21: Comparación del gen PER1 para la variable dicotómica causa de la muerte.

## Gen IDUA

| | CAUSA 1 | | CAUSA 5 | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,64 | 0,948 | 0,804 | 4,423 |
| **COSINOR** | 0,485 | 0,199 | 0,781 | 4,677 |
| **NP** | 0,859 | 0,269 | 0,957 | 4,516 |

Tabla 24: Valores de $R^2$ y peaks para los ajustes del gen IDUA de la variable dicotómica causa de la muerte para los diferentes modelos considerados.

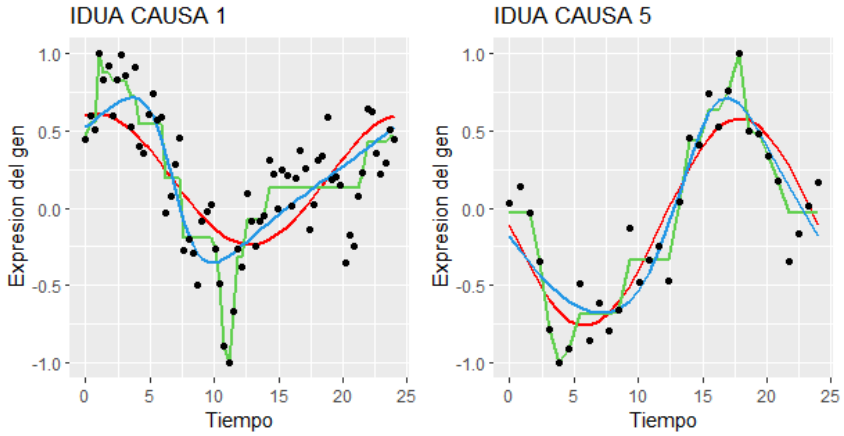


Figura 22: Comparación del gen IDUA para la variable dicotómica causa de la muerte.

## Gen SKI

| | CAUSA 1 | | CAUSA 5 | |
|---|---|---|---|---|
| | R2 | PEAKS | R2 | PEAKS |
| **FMM** | 0,77 | 0,992 | 0,675 | 4,848 |
| **COSINOR** | 0,623 | 0,697 | 0,64 | 4,618 |
| **NP** | 0,887 | 0,808 | 0,91 | 4,516 |

Tabla 25: Valores de $R^2$ y peaks para los ajustes del gen SKI de la variable dicotómica causa de la muerte para los diferentes modelos considerados.

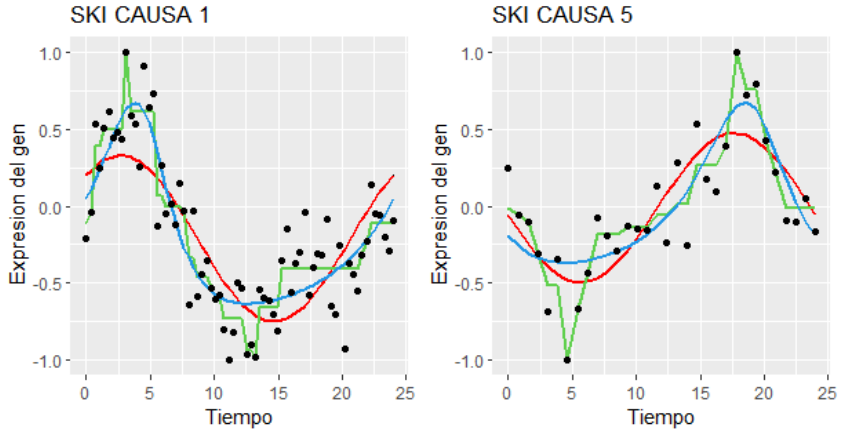


Figura 23: Comparación del gen SKI para la variable dicotómica causa de la muerte.

Para la causa de muerte, gráficamente no se puede afirmar que ambas causas sigan un mismo patrón up-down-up. No obstante, en epígrafes posteriores se llevará a cabo un análisis de las variables dicotómicas que determinará su influencia sobre los patrones rítmicos.

Cabe señalar los genes NR1D1 y PER1 se encuentran entre los que proporcionan mejor ajuste en todas las tablas. El gen NR1D1 es un gen circadiano que regula procesos metabólicos, inflamatorios y cardiovasculares (UnitProt, 2022), mientras que el gen PER1 pertenece a una familia de genes decisivos en la estructura molecular del reloj circadiano en los mamíferos que regula las funciones locomotoras, metabólicas y de comportamiento (GeneCards, 2014).

Los siguientes gráficos permiten valorar el ajuste de cada gen y variable categórica en los distintos modelos. La Tabla 26 contiene los códigos empleados en los ejes de abscisas para representar los genes. Por otro lado, la Tabla 27 contiene los colores empleados para cada categoría.

| GEN | NÚMERO | GEN | NÚMERO |
|---|---|---|---|
| ARNTL | 1 | HLF | 14 |
| NPAS2 | 2 | CRY2 | 15 |
| CLOCK | 3 | PER2 | 16 |
| NFIL3 | 4 | MXD4 | 17 |
| CRY1 | 5 | NELFA | 18 |
| NR1D1 | 6 | FGFRL1 | 19 |
| BHLHE41 | 7 | IDUA | 20 |
| NR1D2 | 8 | ACAP3 | 21 |
| DBP | 9 | CPTP | 22 |
| CIART | 10 | SKI | 23 |
| PER1 | 11 | FAM213B | 24 |
| PER3 | 12 | PRDM16 | 25 |
| TEF | 13 | | |

Tabla 26: Equivalencia genes – números

| CATEGORIA | COLOR |
|---|---|
| Hombre | Negro |
| Mujer | Rojo |
| MSSM | Verde |
| Pitt | Azul |
| Causa 1 | Amarillo |
| Causa 5 | Morado |

Tabla 27: Equivalencia categoría - color

## MODELO FMM

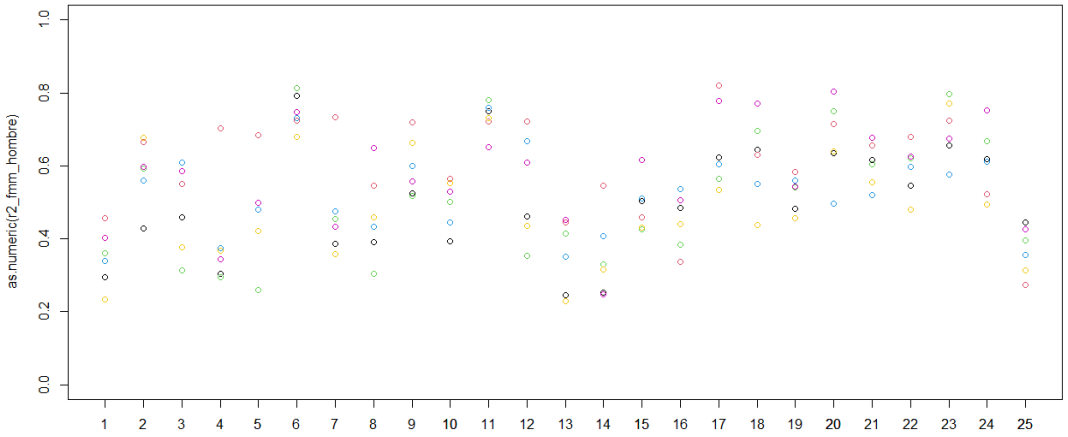


Figura 24: Ajuste modelo FMM para todos los genes y variables

## MODELO COSINOR

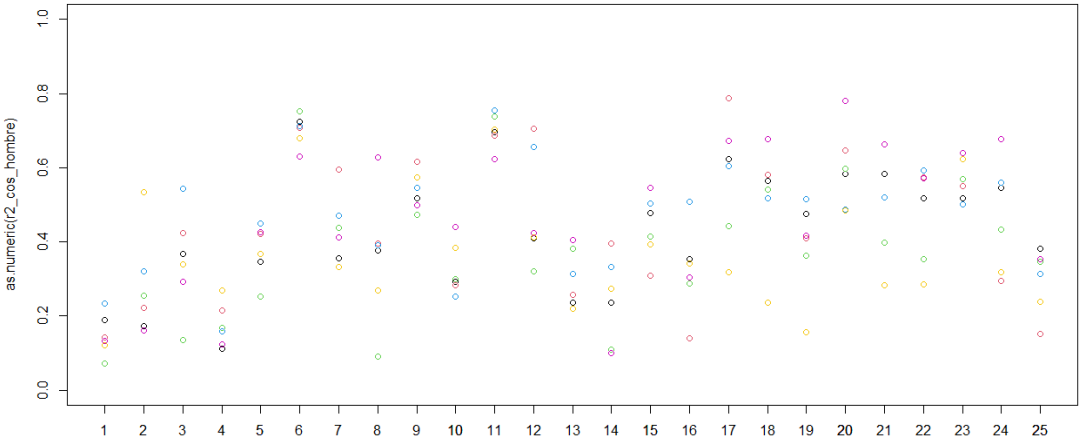


Figura 25: Ajuste modelo Cosinor para todos los genes y variables

## MODELO NP

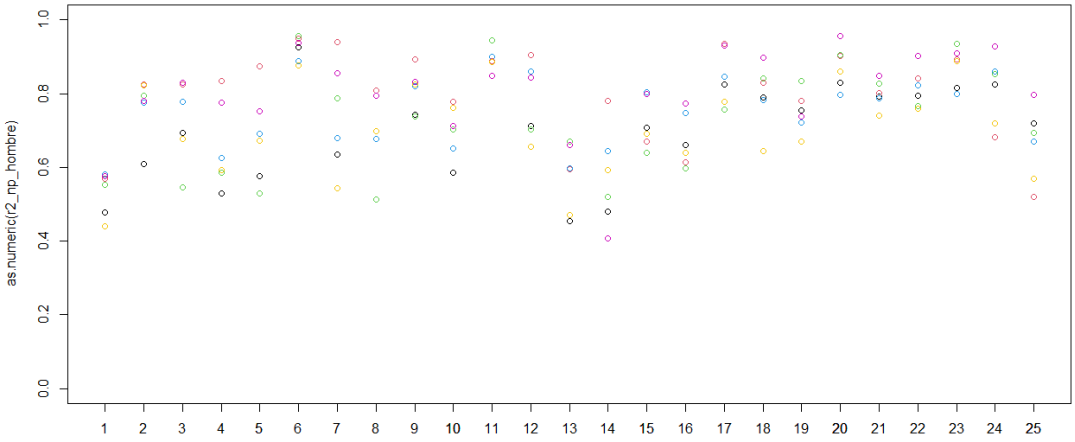


Figura 26: Ajuste modelo NP para todos los genes y variables

Estas figuras representan los resultados anteriores. Globalmente, todos los genes tienen $R^2$ muy elevados y, en particular, los genes NR1D1 (6), PER1 (11), MXD4 (17), NELFA (18), IDUA (20) y SKI (23) presentan los mejores ajustes para cada modelo y variable.

Destacar que el modelo no paramétrico produce mejores resultados que el modelo FMM y el Cosinor, ya que no presupone ninguna distribución de los datos e impone menos restricciones, siendo más flexible y proporcionando mejor ajuste. Además, el modelo Cosinor es un caso particular del modelo FMM, por lo que presenta peor ajuste.

## 4.3 Correlaciones variables dicotómicas

En este epígrafe se estudia la influencia que tienen las variables dicotómicas sobre la configuración de los *peaks*. Para cuantificar su grado de asociación se proponen dos coeficientes de correlación circular: el coeficiente de Jammadamalaka y el de Fisher-Lee, desarrollados en la sección 2.5 de la metodología.

En este caso, tras ajustar los diferentes modelos, los *peaks* se encuentran en el intervalo $[0,2\pi]$, por lo que no es necesario realizar ningún tipo de transformación.

Por otro lado, se parte de dos hipótesis:

- En primer lugar, el origen de los datos no debe influir en la configuración de los *peaks*. Esto implica que los momentos de máxima expresión del gen son independientes de la institución de donde se han recopilado los datos, en este caso, la Universidad de Pittsburgh y la Escuela de medicina Monte Sinaí.
- En segundo lugar, se utiliza como referencia el resultado del modelo Cosinor, ya que es el modelo más empleado en la literatura.

### 4.3.1 Origen de los datos

A continuación, en la Tabla 28 se muestran los resultados del coeficiente de correlación de Fisher-Lee para los *peaks* de cada institución junto con error *bootstrap*.

| MODELO | RESULTADO | ERROR BOOTSTRAP |
|---|---|---|
| COSINOR | 0.516 | 0.168 |
| FMM | 0.139 | 0.149 |
| NP | 0.205 | 0.184 |

Tabla 28: Correlación Fisher-Lee del peak del modelo Cosinor para el origen de los datos

Los resultados de la correlación de Fisher-Lee para el origen de los datos del modelo Cosinor es de 0.5, lo que indica que existe poca asociación entre las instituciones y que éstas podrían influir en los *peaks*. Adicionalmente, el error *bootstrap* obtenido es de 0.168, es decir, el valor obtenido para la correlación esta a 2.88 desviaciones típicas *bootstrap* de 1, siendo este resultado incoherente con la ausencia total de efecto sobre los *peaks.*

Igualmente sucede con los resultados de la correlación del modelo FMM y no paramétrico, cuyos valores se encuentran a más de 4 desviaciones típicas *bootstrap* de 1.

Por ende, se descartan los resultados de este coeficiente ya que no son coherentes la hipótesis de partida.

En la siguiente Tabla 29, se muestran los resultados relativos al coeficiente de correlación de Jammaladamaka junto con el intervalo de confianza al 95% desarrollado en la sección 2.5.1 de la metodología:

| MODELO | RESULTADO | ERROR BOOTSTRAP | INFERIOR | SUPERIOR |
|---|---|---|---|---|
| COSINOR | 0,613 | 0,205 | 0,478 | 1,129 |
| FMM | 0,148 | 0,211 | -0,198 | 0,489 |
| NP | 0,309 | 0,23 | -0,042 | 0,660 |

Tabla 29: Correlación de Jammalamadaka para el origen de los datos

Remarcar que el extremo superior del intervalo como máximo podrá ser 1 puesto que el coeficiente de correlación oscila entre -1 y 1.

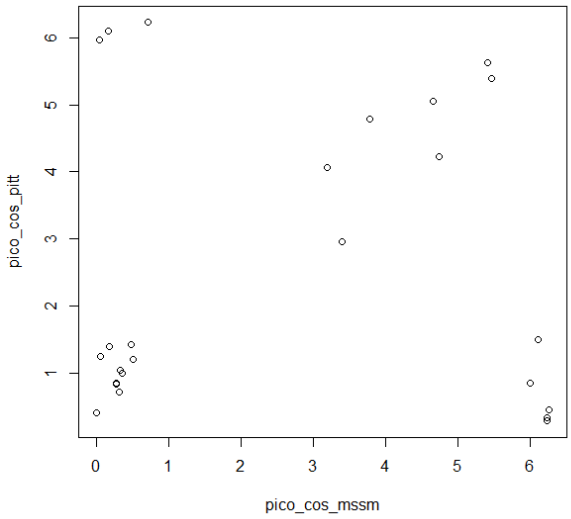


Figura 27: Grafico de correlación para la variable institución

Para el modelo Cosinor se obtiene una correlación de 0.61 que indica asociación. Gráficamente se puede comprobar que ambas variables están correlacionadas de forma positiva. Además, el error boostrap es 0.205 con lo que el valor del estadístico está a 1.89 desviaciones típicas de 1 y, si se considera el intervalo desarrollado en la sección 2.5.1 de este proyecto, su verdadero valor se encuentra entre 0.47 y 1 con una confianza del 95%. Con lo que no se rechazaría el valor 1 a nivel habitual como valor posible para la correlación entre ambas clases, por lo que no existiría influencia en los peaks por parte del origen de los datos. Por consiguiente, en adelante se

utilizará el coeficiente de Jammaladamaka para el análisis, ya que ofrece resultados más razonables.

Los resultados para los modelos FMM y no paramétrico no tienen concordancia respecto a los obtenidos con el Cosinor, esto se debe posiblemente a que ambos modelos son muy sensibles a los *outliers.* Por este motivo, se empleará solo el modelo Cosinor en lo que sigue.

### 4.3.2 Sexo

La Tabla 30 muestra los resultados de la correlación para los *peaks* del sexo junto con su respectivo intervalo de confianza al 95%:

| MODELO | RESULTADO | ERROR BOOTSTRAP | INFERIOR | SUPERIOR |
|---|---|---|---|---|
| COSINOR | 0,927 | 0,035 | 0,768 | 1,339 |

Tabla 30: Correlación de Jammalamadaka e intervalo de confianza de los *peaks* del modelo Cosinor para el sexo

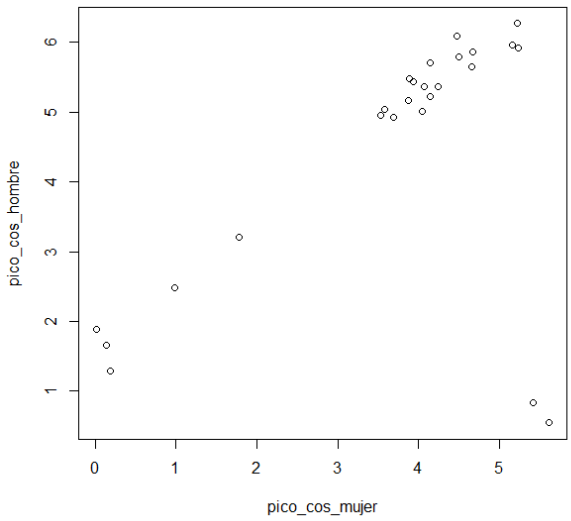


Figura 28: Grafico de correlación para la variable sexo

Empíricamente se obtiene un resultado muy próximo a 1, coherente con la fuerte relación observada gráficamente entre ambos sexos. El error boostrap es 0.035, estando el valor del estadístico a 2 desviaciones típicas *bootstrap* de 1. Asimismo, el verdadero valor del estadístico está comprendido entre 0.76 y 1 con una confianza del 95%, con lo que no se rechazaría el valor 1 a nivel habitual como valor posible para la correlación entre ambos sexos. Como consecuencia, se puede concluir que el sexo no influye significativamente en la configuración de los *peaks*.

### 4.3.3 Causa de la muerte.

Finalmente, la Tabla 31 muestra los resultados para los *peaks* relativos a la causa de la muerte y su intervalo de confianza al 95%.

| MODELO | RESULTADO | ERROR BOOTSTRAP | INFERIOR | SUPERIOR |
|---|---|---|---|---|
| COSINOR | 0,62 | 0,207 | 0,498 | 1,165 |

Tabla 31: Correlación de Jammalamadaka e intervalo de confianza de los *peaks* del modelo Cosinor para la causa de la muerte

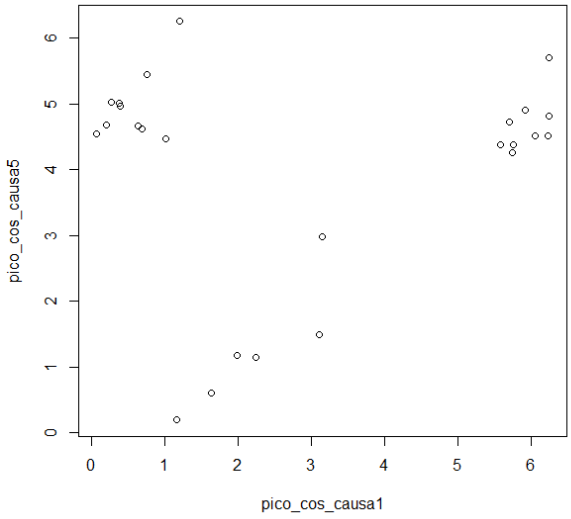


Figura 29: Grafico de correlación para la variable causa de la muerte

Aunque el resultado de la correlación es inferior al del sexo, se considera un resultado alto y se puede observar que existe una relación lineal positiva que permite determinar la existencia de asociación entre la causa 1 y la causa 5. Por otro lado, el valor del estadístico está a 1.9 desviaciones típicas *booststrap* de 1 y su verdadero valor se encontraría entre 0.49 y 1 con una confianza del 95%, consecuentemente, no se rechaza el valor 1 a nivel habitual como posible resultado de la correlación entre ambas causas de fallecimiento y, por tanto, se puede descartar su influencia sobre los *peaks*.

Por ende, se concluye que no existen diferencias significativas entre cada par de variables dicotómicas y que, por tanto, estos factores externos no influyen en los patrones rítmicos de los genes.

### 4.3.4 Edad

Por último, se ha examinado la variable edad, obteniendo resultados muy inestables que no permiten alcanzar una conclusión firme al dividir esta variable en dos grupos, siendo necesaria su estratificación en más categorías. En la Tabla 31 se muestran las correlaciones de los *peaks* del modelo Cosinor para diferentes cortes de edad.

| GRUPOS DE EDAD | CORRELACIÓN | ERROR BOOTSTRAP |
|---|---|---|
| **menores vs mayores 50** | -0,421 | 0,283 |
| **menores vs mayores 55** | 0,180 | 0,324 |
| **menores vs mayores 60** | -0,157 | 0,054 |

Tabla 32: Valores del coeficiente de correlacion de Jammadamalka para diferentes grupos de edad

Sin embargo, aun manteniendo la división en dos categorías, los valores de los estadísticos están a más de 2.5 desviaciones típicas *booststrap* de 1. Por tanto, los resultados obtenidos apuntan a una posible falta de asociación y, consecuentemente, a una influencia efectiva de la variable edad en la configuración de los *peaks*.

Además, cabe resaltar que algunos estudios señalan la tendencia de las personas mayores a despertarse temprano y acostarse pronto. Curiosamente, existe una afección llamada “Síndrome de la puesta del sol”, que afecta al 20%-40% de las personas mayores con demencia o Alzheimer. Esta enfermedad hace que las personas que lo padecen se sientan confusas, ansiosas e incluso delirantes ante la puesta del sol. Esto provoca que deambulen y tengan dificultades para conciliar el sueño (Chen et al., 2015).

Por lo tanto, se cree que la puesta de sol podría estar directamente relacionada con la ruptura de los ritmos circadianos en estos individuos. Aunque en la muestra de datos recogida solo se consideran individuos sanos, es posible que algunos de los genes que experimentan una pérdida de ritmo con el envejecimiento puedan estar relacionados con cambios en la cognición (Chen et al., 2015).

# 5. DISCUSIÓN, CONCLUSIONES Y TRABAJO FUTURO

## 5.1 Discusión y conclusiones

El objetivo de este proyecto era el estudio de los factores influyentes en la ritmicidad de expresiones de genes a partir de datos post mortem. Para ello, se ha desarrollado una metodología que valora la influencia de variables dicotómicas en los momentos de máxima expresión de los genes rítmicos, instante en el que ejercen su función en el ciclo celular. Con el fin de asegurar la ritmicidad de los genes considerados se ha establecido un conjunto de genes *core* para los que la ritmicidad ya ha sido valorada en otros estudios.

Para el desarrollo de la metodología se han considerado dos modelos alternativos al modelo Cosinor, que es el más empleado en la literatura, el modelo FMM y el no paramétrico. Cabe destacar que estos dos últimos aportan la suficiente flexibilidad para adaptarse a una gran variedad de patrones rítmicos, entre los que se incluyen las formas asimétricas.

Para reflejar la bondad de ajuste de los modelos, se ha empleado el coeficiente de determinación $R^2$. Los genes: NR1D1, PER1, MXD4, NELFA, IDUA y SKI, cuyos $R^2$ han oscilado entre 0.6 y 0.8, han presentado los mejores ajustes para cada modelo y variable.

Además, se han considerado dos coeficientes de correlación circular para estudiar la asociación entre los factores. Por un lado, el coeficiente de Fisher-Lee, que ha sido descartado, junto al modelo FMM y no paramétrico por arrojar resultados incompatibles con la hipótesis inicial de asociación entre los datos recogidos por la Universidad de Pittsburg y la Escuela de Medicina de Monte Sinaí. Por otro, el coeficiente de correlación de Jammalamadaka que, junto al modelo Cosinor, ha producido resultados más estables, permitiendo descartar la influencia de las variables sexo y causa de la muerte sobre los momentos de máxima expresión de los genes *core* seleccionados. Concretamente, los *peaks* de los hombres y las mujeres han presentado una fuerte asociación positiva con una correlación de 0.92. La correlación de *peaks* de las causas de fallecimiento ha sido de 0.62 pero no se ha rechazado el valor 1 como posible resultado de la correlación.

Tras el análisis expuesto, se puede concluir que no se han encontrado diferencias significativas entre cada par de variables dicotómicas, por lo que estos factores externos podrían no influir en los patrones rítmicos de los genes.

También se ha diseñado y programado en R el intervalo de confianza para el coeficiente de la correlación de Jammalamadaka, que permite aproximar el rango de valores entre los que se encuentra su verdadero valor.

## 5.2 Trabajo futuro

Este proyecto ha abierto varias líneas de investigación futuras que pueden agruparse en dos tipos generales de trabajo.

Trabajo metodológico

- Las diferencias observadas entre los coeficientes de Jammalamadaka y Fisher-Lee, aparentemente similares en su definición, a la hora de valorar las asociaciones entre los momentos de máxima expresión de los genes hacen que sea interesante un estudio comparativo tanto desde el punto de vista teórico como de simulación entre ambos coeficientes para determinar el origen de estas diferencias. No se ha encontrado tal estudio en la literatura existente hasta el momento.

- Llama la atención también la sensibilidad que ha mostrado el modelo FMM a los valores extremos, posiblemente atípicos, que con frecuencia aparecen en los datos como los tratados en este trabajo. Si bien esto posiblemente esté relacionado con la flexibilidad que a este modelo le confieren sus parámetros, parece conveniente un estudio detallado de hasta qué punto estas observaciones influyen en los resultados que se obtienen a partir de este modelo.

- En este trabajo se ha considerado también la variable edad categorizándola. Sin embargo, los resultados obtenidos no son lo suficientemente estables como para extraer conclusiones. Por lo observado en el trabajo realizado por esta variable parece conveniente un estudio de la robustez de los resultados que se obtienen a partir del método de estimación de órdenes ya que estos parecen depender en exceso de los valores de arranque del algoritmo TSP subyacente.

Trabajo aplicado

- Se han considerado aquí variables dicotómicas. Sería interesante el diseño de métodos válidos para variables cualitativas con más de dos categorías. Una posibilidad es la comparación de los órdenes obtenidos con cada una de las categorías con el orden global. Para ello, se sugiere una extensión del procedimiento ORI que permita obtener órdenes temporales no equiespaciados. De esta forma, se podría estudiar la correlación entre la ordenación obtenida.

- También queda pendiente el estudio de la influencia de variables continuas sobre los momentos de máxima expresión. Hay que notar que la variable edad entra dentro de esta situación si no se categorizan sus valores.

# BIBLIOGRAFÍA


Anafi, R., Francey, L., Hogenesch, J., & Junhyong, K. (2017). CYCLOPS reveals human transcriptional rhythms in health and disease. *Proceedings of the National Academy of Sciences of the United States of America*, 114(20): 5312-5317. doi:https://doi.org/10.1073/pnas.1619320114

Borja, M. (Noviembre de 2019). *Ritmos circadianos, qué son y cómo influyen en nuestra salud*. Obtenido de Salud: https://www.20minutos.es/noticia/4062540/0/ritmos-circadianos-que-son-y-como-influyen-en-la-salud/

Cano, V. (Abril de 2018). *Qué son los ritmos circadianos y cómo debes cuidar el tuyo*. Obtenido de Business Insider: https://www.nigms.nih.gov/education/fact-sheets/Pages/circadian-rhythms-spanish.aspx

Chen, C.-Y., Wellington, R., Ma, T., & Mcclung, C. (2015). Effects of aging on circadian patterns of gene expresssion in the human prefrontal cortex. *CrossMark*, 113(1): 206-211. doi:https://doi/abs/10.1073/pnas.1508249112

Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling*, 11 Article number:16. doi:10.1186/1742-4682-11-16

Fisher, R., & Lee, E. (1995). Circular-circular association: T-linear association. En N. I. Fisher, *Statistical Analysis Of Circular Data* (págs. 327-330). Great Britain: Cambridge University Press.

GeneCards. (Marzo de 2014). *PER1 Gene - Period Circadian Regulator 1*. Obtenido de NCBI: https://www.ncbi.nlm.nih.gov/gene/5187

Griffiths, A., & Lewontin, W. (2004). *Introduction to Genetic Analysis.* Londres: W.H.Freeman & Co Ltd.

Jammalamadaka, S., & Sarma, R. (1988). A correlation coefficient for angular variables. En S. R. Jammalamadaka, & Sarma, *Statistical Theory and Data Analysis II* (págs. 349-364). Holland: Elsevier Science Publishers.

Larriba, Y., Rueda, C., Fernández , M., & Peddada, S. (2020). Order restricted inference in chronobiology. *Statistics in Medicine*, 39(3): 265-278. doi:10.1002/sim.8397

Leng, N., Chu, L.-F., Chris, B., Yuan, L., Jeea, C., Xiamo, L., . . . Kendziorsk, C. (2015). Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nature Methods*, 12(10): 947-950. doi:10.1038/nmeth.3549

Li, J., Bunney, B., Meng, F., & Bunney, W. (2013). Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proceedings of the National Academy of Science*, 110(24): 9950-9955. doi:https://doi.org/10.1073/pnas.1305814110

Liu, C., Gershon, E., & Kelsoe, J. (2017). From Gene Expression To Disease Association. *European Neuropsychopharmacology*, 27(3): 416-463. doi:10.1016/j.euroneuro.2016.09.463

Lorsch, J. (Enero de 2021). *National Institute of General Medical Sciences*. Obtenido de Services, U.S. Department of Health and Human: https://www.nigms.nih.gov/education/fact-sheets/Pages/circadian-rhythms-spanish.aspx

Mardia, K., & Jupp, G. (2000). Statistics of directional data. En Mardia, *Directional Statistics* (págs. 349-393). Wiley.

Mendel, V. (Junio de 2017). *El insomnio y los Zeitgebers.* Obtenido de Carcaj flechas de sentido: http://carcaj.cl/el-insomnio-y-los-zeitgebers/

Mure, L. S., Le, H. D., Benegiamo, G., Chang, M. W., Rios, L., Jillani, N., & Ngotho, M. (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *National Center for Biotechnology Information*, 359(6381): Issue 6381. doi:10.1126/science.aao0318

Piacente, P. J. (Marzo de 2021). *Tendencias*. Obtenido de Tendencias: https://tendencias21.levante-emv.com/hay-genes-en-el-cerebro-que-sobreviven-a-la-muerte.html

Prieto, C. M. (2021). *Una propuesta novedosa para la estimación de la hora de la muerte a partir de datos post mortem de expresiones de genes.* Universidad de Valladolid: Trabajo de Fin de Grado.

Robertson, T., Dykstra, R., & Wright, F. (1988). Order Restricted Statistical Inference. Wiley.

Ruben, M. D., Wu, G., Smith, D. F., & Schmidt, R. E. (2018). A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Science Translational Medicine*, 10(458): eaat8806. doi:10.1126/scitranslmed.aat8806

Rueda, C., Larriba, Y., & Peddada, S. D. (2019). Frequency Modulated Möbius Model Accurately Predicts Rhythmic Signals in Biological and Physical Sciences. *Scientific Reports*, 9, Article number: 18701. doi:10.1038/s41598-019-54569-1

Seney, M., Cahill, K., Enwright, J., Logan, R., Hou, Z., Tseng, G., & Mcclung, C. (2019). Diurnal rhythms in gene expression in the prefrontal cortex in

schizophrenia. *Nature Communications*, 10(1): 1-11. doi:10.1038/s41467-019-11335-1

UnitProt. (Junio de 2022). *Nr1d1 nuclear receptor subfamily 1, group D, member 1 [ Mus musculus (house mouse) ]*. Obtenido de Unit Pro: https://www.uniprot.org/uniprot/Q3UV55

# ANEXO A: LECTURA DE DATOS, GENERACIÓN DE ÓRDENES Y SINCRONIZACIÓN

```
#################################### LECTURA DE DATOS
####################################
#Fichero de individuos y genes Filtered_log2CPM.csv
inData<-read.csv(file=paste(ruta,"Filtered_log2CPM.csv",sep=""), header=TRUE,
sep=",")[,-1]
inData<-as.matrix(inData)
rownames(inData)<-read.csv(file=paste(ruta,"Filtered_log2CPM.csv",sep=""), header=TRUE,
sep=",")[,1]

# Fichero con la hora de muerte data1Seney2019.xlsx
#deathData <- read.xlsx(file=paste(ruta,"data1Seney2019.xlsx",sep=""), 1,header=TRUE,
sep=",")
deathData <-read_excel("TFG/data1Seney2019.xlsx")

#Fichero con todas las caracteristicas CMC_MSSM-Penn-Pitt_Clinicalv2.csv
clinicalData<-read.csv(file=paste(ruta,"CMC_MSSM-Penn-Pitt_Clinicalv2.csv",sep=""),
header=TRUE, sep=";")#[,1]
clinicalData<-as.matrix(clinicalData)
controlDeath<- subset(deathData, Dx == "Control")
controlData<-merge(controlDeath, clinicalData, by="Individual_ID")

#Devuelve los indices
indControlSubjects<-
match(controlData[,"DLPFC_RNA_Sequencing_Sample_ID"],colnames(inData))
dataControlSubjects<-inData[,indControlSubjects]
controlData<- as.matrix(controlData)


controlData <- as.data.frame(controlData)
controlData <- controlData %>% mutate(Age_cat = ifelse(Age_of_Death < 55,1,2))
controlData <- controlData %>% mutate(PMI_cat = ifelse(PMI_hrs < 12, 1, ifelse(PMI_hrs <
18 & PMI_hrs >= 12, 2, ifelse(PMI_hrs < 24 & PMI_hrs >= 18,3,4))))
#################################### OBTENER ORDEN ORI
####################################
seleccion_parametros_orden <- function(ori){

  vector <- ori$DLPFC_RNA_Sequencing_Sample_ID

  #filtramos los genes core y las mujeres
  geneNames <-
c("ARNTL","NPAS2","CLOCK","NFIL3","CRY1","NR1D1","BHLHE41","NR1D2","DBP","CIART","PER1",
"PER3","TEF","HLF","CRY2","PER2","MXD4","NELFA","FGFRL1","IDUA","ACAP3","CPTP","SKI","FA
M213B","PRDM16")
  geneSelectionData <- dataControlSubjects[geneNames,]
  copia_geneSelectionData <- geneSelectionData
  geneSelectionData <- geneSelectionData[,vector]
  nombres <- colnames(geneSelectionData)

  ORI_order_Reduced<-TSP_Euc_v7(datos=geneSelectionData,dist_type="Euclidea",
                                datos2p=dataControlSubjects,pesosE=FALSE,pesosMSE=TRUE,
                                pesosAdd=FALSE,unPeriodo=TRUE,
                               pesos=rep(1,ncol(gn)),
                          centrar=FALSE,intentos=15,onlyHeuristica2=FALSE)
 print(ORI_order_Reduced)
  h <- ORI_order_Reduced[[1]]
  return(h)
}



#################################### OBTENER ORDEN EQUIVALENCIA MUESTRA TOTAL
####################################
ori <- controlData %>% filter(Institution.x == "Pitt")
vector <- ori$DLPFC_RNA_Sequencing_Sample_ID

#filtramos los genes core y las mujeres
```

```
geneNames <-
c("ARNTL","NPAS2","CLOCK","NFIL3","CRY1","NR1D1","BHLHE41","NR1D2","DBP","CIART","PER1",
"PER3","TEF","HLF","CRY2","PER2","MXD4","NELFA","FGFRL1","IDUA","ACAP3","CPTP","SKI","FA
M213B","PRDM16")
geneSelectionData <- dataControlSubjects[geneNames,]
copia_geneSelectionData <- geneSelectionData
geneSelectionData <- geneSelectionData[,vector]
nombres <- colnames(geneSelectionData)

posicion <- c()
#obtener la posicion
for(i in 1:dim(ori)[1]){
  posicion[i] <- which(colnames(copia_geneSelectionData) == nombres[i])

}
posicion <- sort(posicion)
#posicion


orden <- orden_todos[orden_todos%in% posicion]

#obtener la equivalencia


equivalencia <- function(orden){
  dim <-length(orden)
  aux <- orden
  equivalencia <- aux
  contador <- 1
  for(i in 1:dim){
    equivalencia[which.min(aux)] <- contador
    aux[which.min(aux)] = 1000
    contador = contador +1
  }
  print(equivalencia)
}
equivalencia(orden)

######################################## SINCRONIZACIÓN DE GENES Y ORDENES
########################################
#obtener el más alto
centrar <-function(indORIDataReduced,ori,time)  {
  geneNames <-
c("ARNTL","NPAS2","CLOCK","NFIL3","CRY1","NR1D1","BHLHE41","NR1D2","DBP","CIART","PER1",
"PER3","TEF","HLF","CRY2","PER2","MXD4","NELFA","FGFRL1","IDUA","ACAP3","CPTP","SKI","FA
M213B","PRDM16")
  data <- indORIDataReduced
  nGen <- nrow(indORIDataReduced)
  nIndv <-dim(ori)[1]
  time <- time
  period <- 24

  picos_NP_R2 <- matrix(nrow=nGen, ncol = 2)

  for (i in 1:nGen){#nrow(indORIDataReducedSCZ)){
    data[i,]<- rescale(data[i,], to=c(-1,1))
    NP<- function1Local(data[i,])
    picos_NP_R2[i,]<-c(rownames(data)[i],NP[[6]])
  }
  arntl <- as.numeric(picos_NP_R2[1,2])
  x <- dim(indORIDataReduced)[[2]]
  x <- round(x/2)
  if( arntl>x){
    pos <- arntl - x
    #leng <- dim(geneSelectionData)[[2]] - pos
    vec1 <- indORIDataReduced[,1:pos]
    vec2 <- indORIDataReduced[,-c(1:pos)]
    vector <- cbind(vec2,vec1)
  }
  if( arntl<x){
    pos <- x - arntl
```

```
    #leng <- dim(geneSelectionData)[[2]] - pos
    vec1 <- indORIDataReduced[,(dim(indORIDataReduced)[2]-
pos):(dim(indORIDataReduced)[2])]
    vec2 <- indORIDataReduced[,-c((dim(indORIDataReduced)[2]-
pos):(dim(indORIDataReduced)[2]))]
    vector <- cbind(vec1,vec2)
  }
  if( arntl == x){
    vector <- indORIDataReduced
  }
  return(vector)
}

centrar_orden <-function(indORIDataReduced,ori,time,orden)  {
  geneNames <-
c("ARNTL","NPAS2","CLOCK","NFIL3","CRY1","NR1D1","BHLHE41","NR1D2","DBP","CIART","PER1",
"PER3","TEF","HLF","CRY2","PER2","MXD4","NELFA","FGFRL1","IDUA","ACAP3","CPTP","SKI","FA
M213B","PRDM16")
  data <- indORIDataReduced
  nGen <- nrow(indORIDataReduced)
  nIndv <-dim(ori)[1]
  time <- time
  period <- 24
  picos_NP_R2 <- matrix(nrow=nGen, ncol = 2)

  for (i in 1:nGen){#nrow(indORIDataReducedSCZ)){
    data[i,]<- rescale(data[i,], to=c(-1,1))
    NP<- function1Local(data[i,])
    picos_NP_R2[i,]<-c(rownames(data)[i],NP[[6]])
  }
  arntl <- as.numeric(picos_NP_R2[1,2])
  x <- dim(indORIDataReduced)[[2]]
  x <- round(x/2)
  if(arntl>x){
    pos <- arntl - x
    #leng <- dim(geneSelectionData)[[2]] - pos
    vec1 <- orden[1:pos]
    vec2 <- orden[-c(1:pos)]
    vector <- c(vec2,vec1)
  }
  if( arntl<x){
    pos <- x - arntl
    #leng <- dim(geneSelectionData)[[2]] - pos
    vec1 <- orden[(length(orden)-pos):(length(orden))]
    vec2 <- orden[-c((length(orden)-pos):(length(orden)))]
    vector <- c(vec1,vec2)
  }
  if( arntl == x){
    vector <- orden
  }
  return(vector)
}
```

# ANEXO B: AJUSTE DE MODELOS Y CORRELACIONES

```
####OBTENCION R2 Y PICOS####

#ANADIR GENES
seleccion_parametros_orden <- function(ori){

  vector <- ori$DLPFC_RNA_Sequencing_Sample_ID

  #filtramos los genes core y las mujeres
  geneNames <-
c("ARNTL","NPAS2","CLOCK","NFIL3","CRY1","NR1D1","BHLHE41","NR1D2","DBP","CIART","PER1",
"PER3","TEF","HLF","CRY2","PER2","MXD4","NELFA","FGFRL1","IDUA","ACAP3","CPTP","SKI","FA
M213B","PRDM16")
  geneSelectionData <- dataControlSubjects[geneNames,]
  copia_geneSelectionData <- geneSelectionData
  geneSelectionData <- geneSelectionData[,vector]
  nombres <- colnames(geneSelectionData)

  return(geneSelectionData)
}


# R2 Y PEAKS
r2_picos <- function(ori,indORIDataReduced){ #(ori,indORIDataReduced,geneSelectionData
  time<-rescale(rep(1:dim(ori)[1]),to=c(0, 2 * pi))
  periodo<-24

  errorORIReduced <- calculoError(indORIDataReduced,nrow(indORIDataReduced),dim(ori)[1],
time, 24)#, "NP_ORI_Reducido_Control", "Cos_ORI_Reducido_Control",
"FMM_ORI_Reducido_Control")

  colnames(errorORIReduced$error)<-
c("gen","ORI_Reducido_Control_NP_Error","ORI_Reducido_Control_Cos_Error",
"ORI_Reducido_Control_FMM_Error")
  colnames(errorORIReduced$R2)<-
c("gen","ORI_Reducido_Control_NP_R2","ORI_Reducido_Control_Cos_R2",
"ORI_Reducido_Control_FMM_R2")
  colnames(errorORIReduced$parametros_Cosinor)<-
c("gen","ORI_Reducido_Control_Cosinor_M","ORI_Reducido_Control_Cosinor_A",
"ORI_Reducido_Control_Cosinor_phi")
  colnames(errorORIReduced$parametros_FMM)<-
c("gen","ORI_Reducido_Control_FMM_M","ORI_Reducido_Control_FMM_A",
"ORI_Reducido_Control_FMM_alpha","ORI_Reducido_Control_FMM_beta","ORI_Reducido_Control_F
MM_omega")
  colnames(errorORIReduced$pico_Cosinor)<-
c("gen","ORI_Reducido_Control_Cosinor_ZL","ORI_Reducido_Control_Cosinor_ZU",
"ORI_Reducido_Control_Cosinor_TL", "ORI_Reducido_Control_Cosinor_TU")
  colnames(errorORIReduced$pico_FMM)<-
c("gen","ORI_Reducido_Control_FMM_ZL","ORI_Reducido_Control_FMM_ZU",
"ORI_Reducido_Control_FMM_TL", "ORI_Reducido_Control_FMM_TU")



  error_R2 <- as.data.frame(errorORIReduced)
  pico <- error_R2%>%
select(error.gen,pico_FMM.ORI_Reducido_Control_FMM_ZL,pico_FMM.ORI_Reducido_Control_FMM_
TU,pico_FMM.ORI_Reducido_Control_FMM_TL,pico_FMM.ORI_Reducido_Control_FMM_ZU)
  tablaR2 <- error_R2 %>%
select(R2.gen,R2.ORI_Reducido_Control_NP_R2,R2.ORI_Reducido_Control_Cos_R2,R2.ORI_Reduci
do_Control_FMM_R2)
  return(error_R2)
}

#PEAKS MODELO NP
peaks_NP <- function(indORIDataReduced,ori,time,orden1){
  geneNames <-
c("ARNTL","NPAS2","CLOCK","NFIL3","CRY1","NR1D1","BHLHE41","NR1D2","DBP","CIART","PER1",
"PER3","TEF","HLF","CRY2","PER2","MXD4","NELFA","FGFRL1","IDUA","ACAP3","CPTP","SKI","FA
M213B","PRDM16")
  data <- indORIDataReduced
```

```
  nGen <- nrow(indORIDataReduced)
  nIndv <-dim(ori)[1]
  time <- time
  period <- 24

  picos_NP_R2 <- matrix(nrow=nGen, ncol = 2)

  for (i in 1:nGen){#nrow(indORIDataReducedSCZ)){
    data[i,]<- rescale(data[i,], to=c(-1,1))
    NP<- function1Local(data[i,])
    picos_NP_R2[i,]<-c(rownames(data)[i],NP[[6]])
  }

  seq1<-seq(0,2*pi,length.out=length(orden1)+1)
  seq1<-seq1[1:length(orden1)]

  peaks <- matrix(nrow=nGen, ncol = 2)
  picos_NP_R2 <- as.data.frame(picos_NP_R2)
  picos_NP_R2$V2 <- as.numeric(picos_NP_R2$V2)
  for (i in 1:nGen){
    data[i,]<- rescale(data[i,], to=c(-1,1))
    peaks[i,] <-c(rownames(data)[i],seq1[picos_NP_R2$V2[i]])
  }
  return(peaks)

}
######################################## OBTENER CORRELACIONES ORDENES
########################################
estimadores <- function(orden1,orden2){
  seq1<-seq(0,2*pi,length.out=length(orden1)+1)
  seq1<-seq1[1:length(orden1)]

  no1<-order(orden1)
  no2<-order(orden2)

  phi1<-seq1[no1]
  phi2<-seq1[no2]

  phih <- cbind(phi1,phi2)

  resultado <- circ_cor(phih,type="js", alternative = "greater",bootse = TRUE,n.boot =
1000)
  js <- resultado[1]
  se <- attr(resultado,"se")
  bootse <- attr(resultado,"bootse")
  datos <- cbind(resultado,se,bootse)
  return(datos)
}
######################################## OBTENER CORRELACIONES PEAKS
########################################
estimadores <- function(phi1,phi2){
  phi1 <- as.numeric(phi1)
  phi2 <- as.numeric(phi2)
  phih <- cbind(phi1,phi2)

  resultado <- circ_cor(phih,type="tau1", alternative = "greater",bootse = TRUE,n.boot =
1000)
  js <- resultado[1]
  se <- attr(resultado,"se")
  bootse <- attr(resultado,"bootse")
  datos <- cbind(resultado,se,bootse)
  return(datos)
}
######################################## INTERVALO DE CONFIANZA Y CORRELACION DE
JAMMALAMADAKA ############ ############################
intervalo <- function(phi1,phi2){

  phi1 <- as.numeric(phi1)
  phi2 <- as.numeric(phi2)

  phih <- cbind(phi1,phi2)
```

```
  #circ_cor(phih,type="js",alternative = "greater")
  rho <- circ_cor(phih,type="js",alternative = "greater")[1]

  est <- sqrt(nrow(phih))*(rho - 1)
  zphi1 <- phi1 - 3.1415
  zphi2 <- phi2 - 3.1415
  media_phi1 <- mean.circular(zphi1)[[1]]
  media_phi2 <- mean.circular(zphi2)[[1]]
  landa_est <- function(i,j){
    resultado <- 0
    for(h in 1:nrow(phih)){
      resultado <- (sin(phi1[h]-media_phi1)^i)*(sin(phi2[h]-media_phi2)^j) + resultado
    }
    return((1/nrow(phih))*resultado)
  }

  var           <-           (landa_est(2,2)/(landa_est(2,0)*landa_est(0,2)))           -
rho*(landa_est(1,3)/(landa_est(2,0)*sqrt(landa_est(2,0)*landa_est(0,2)))+landa_est(3,1)/
(landa_est(0,2)*sqrt(landa_est(2,0)*landa_est(0,2))))
  var                           <-                          var                           +
(rho^2/4)*(1+(landa_est(4,0)/(landa_est(2,0)^2))+(landa_est(0,4)/(landa_est(0,2)^2))+(la
nda_est(2,2)/(landa_est(2,0)*landa_est(0,2))))

  estadistico <- est/sqrt(var)

  #p(t< -7.29) ; t~n(0,1)
  pval <- pnorm(estadistico)

  low <- (rho - 1.96*sqrt(var)/sqrt(nrow(phih)))/sqrt(nrow(phih))
  high <- (rho + 1.96*sqrt(var)/sqrt(nrow(phih)))/sqrt(nrow(phih))
  result<-c(rho,estadistico,pval,low,high)


  return(result)


}
```

# ANEXO C: FUNCIONES Y LIBRERÍAS

```
library(readxl)
library(tidyverse)
library(dplyr)
library(FMM)
library(TSP)
library(Iso)
library(foreach)
library(iterators)
library(parallel)
library(doParallel)
library(bigstatsr)
library(scales)
library(xlsx)
library(ggplot2)
library(cosinor2)
library(gridExtra)
library(grid)
library(circular)
library("writexl")
library("ggplotify")
library(isocir)
library(BAMBI)
library(Rcpp)
library(ggpubr)
library(plyr)
library(reshape2)
library(grid)


########FUNCIONES

function1Local<-function(v){
  require(Iso)
  v2<-c(v,v)
  i2<-c(1:length(v),1:length(v))
  candL<-c()
  candU<-c()
  for(i in (2):(length(v)+1)){
    #maximos locales
    if(v2[i-1]<=v2[i] & v2[i]>=v2[i+1]){
      candiU<-i2[i]
      candU<-unique(c(candU,candiU))
    }

    #minimos locales
    if(v2[i-1]>=v2[i] & v2[i]<=v2[i+1]){
      candiL<-i2[i]
      candL<-unique(c(candL,candiL))
    }

  }


  mseFin<-9999999999999999999999
  for( i in 1:length(candL) ){
    for (j in 1:length(candU) ){
      #print(c(i,j))
      #i<-210
      #j<-168
      if(candL[i]<candU[j]){
        indexPavaLU<-candL[i]:candU[j]
        if(length(indexPavaLU)>2){
          pavaLU<-pava(v[(candL[i]+1):(candU[j]-1)])
          pavaLU<-c(v[candL[i]],pavaLU,v[candU[j]])
        }else{
          pavaLU<-c(v[candL[i]],v[candU[j]])
        }
        if(candU[j]!=length(v) & candL[i]!=1){
          indexPavaUL1<-(candU[j]+1):length(v)
```

```
          indexPavaUL2<-1:(candL[i]-1)
          pavaUL<-pava(v2[(candU[j]+1):(length(v)+candL[i]-1)],decreasing=TRUE)
          pavaAux<-
c(pavaUL[(length(indexPavaUL1)+1):length(pavaUL)],pavaLU,pavaUL[1:length(indexPavaUL1)])
          if((length(v)-length(pavaAux))!=0){
            #print("##########################################################error")
          }
          mseAux<-sum((v-pavaAux)^2)/length(v)
          if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
             pavaUL[1]<=v[candU[j]] & pavaUL[length(pavaUL)]>=v[candL[i]] &
mseAux<mseFin){#=
            mseFin<-mseAux
            pavaFin<-pavaAux
            Lopt<-candL[i]
            Uopt<-candU[j]
          }
        }
        if(candU[j]==length(v) & candL[i]!=1){
          pavaUL<-pava(v[1:(candL[i]-1)],decreasing=TRUE)
          pavaAux<-c(pavaUL,pavaLU)
          if((length(v)-length(pavaAux))!=0){
            #print("##########################################################error")
          }
          mseAux<-sum((v-pavaAux)^2)/length(v)
          if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
             pavaUL[1]<=v[candU[j]] & pavaUL[length(pavaUL)]>=v[candL[i]] &
mseAux<mseFin){
            mseFin<-mseAux
            pavaFin<-pavaAux
            Lopt<-candL[i]
            Uopt<-candU[j]
          }
        }
        if(candL[i]==1 & candU[j]!=length(v)){
          pavaUL<-pava(v[(candU[j]+1):(length(v))],decreasing=TRUE)
          pavaAux<-c(pavaLU,pavaUL)
          if((length(v)-length(pavaAux))!=0){
            #print("##########################################################error")
          }
          mseAux<-sum((v-pavaAux)^2)/length(v)
          if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
             pavaUL[1]<=v[candU[j]] & pavaUL[length(pavaUL)]>=v[candL[i]] &
mseAux<mseFin){
            mseFin<-mseAux
            pavaFin<-pavaAux
            Lopt<-candL[i]
            Uopt<-candU[j]
          }
        }
        if(candL[i]==1 & candU[j]==length(v)){
          pavaUL<-c()
          pavaAux<-pavaLU
          if((length(v)-length(pavaAux))!=0){
            #print("##########################################################error")
          }
          mseAux<-sum((v-pavaAux)^2)/length(v)
          if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
mseAux<mseFin){
            mseFin<-mseAux
            pavaFin<-pavaAux
            Lopt<-candL[i]
            Uopt<-candU[j]
          }
        }
      }else{#candU<=candL#revisar
        #agnado un if oara cuando son inguales los candidatos el else es lo que habia
para candU<candL
        if(candL[i]==candU[j]){
          pavaAux<-rep(mean(v),length(v))
          mseAux<-sum((v-pavaAux)^2)/length(v)
          #if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
```

```
          #   pavaUL[1]<v[candU[j]] & pavaUL[length(pavaUL)]>v[candL[i]] &
mseAux<mseFin){
          if( mseAux<mseFin){
            mseFin<-mseAux
            pavaFin<-pavaAux
            Lopt<-candL[i]
            Uopt<-candU[j]
          }
        }else{
          indexPavaLU<-c(candL[i]:length(v),1:candU[j])
          if(length(indexPavaLU)>2){
            pavaLU<-pava(v2[(candL[i]+1):(length(v)+candU[j]-1)])
            pavaLU<-c(v[candL[i]],pavaLU,v[candU[j]])
          }else{
            pavaLU<-c(v[candL[i]],v[candU[j]])
          }
          if(length(pavaLU)==length(v)){
            pavaAux<-
c(pavaLU[(length(candL[i]:length(v))+1):length(pavaLU)],pavaLU[1:length(candL[i]:length(
v))])
            if((length(v)-length(pavaAux))!=0){
              #print("##########################################################error")
            }
            mseAux<-sum((v-pavaAux)^2)/length(v)
            if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]]  &
mseAux<mseFin){
              mseFin<-mseAux
              pavaFin<-pavaAux
              Lopt<-candL[i]
              Uopt<-candU[j]
            }
          }else{
            if(candU[j]!=1 & candL[i]!=length(v)){
              #if(length(pavaLU)!=length(v)){
              pavaUL<-pava(v[(candU[j]+1):(candL[i]-1)],decreasing=TRUE)
              pavaAux<-
c(pavaLU[(length(candL[i]:length(v))+1):length(pavaLU)],pavaUL,pavaLU[1:length(candL[i]:
length(v))])
              if((length(v)-length(pavaAux))!=0){
                # print("##########################################################error")
              }
              mseAux<-sum((v-pavaAux)^2)/length(v)
              if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
                 pavaUL[1]<=v[candU[j]] & pavaUL[length(pavaUL)]>=v[candL[i]] &
mseAux<mseFin){
                mseFin<-mseAux
                pavaFin<-pavaAux
                Lopt<-candL[i]
                Uopt<-candU[j]
              }
              #}else{
              #pavaUL<-c()
              #pavaAux<-
pavaLU[c((length(candL[i]:length(v))+1):(length(v)),1:(length(candL[i]:length(v))))]
              #mseAux<-sum((v-pavaAux)^2)/length(v)
              #if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]]  &
mseAux<mseFin){
              #mseFin<-mseAux
              #pavaFin<-pavaAux
              #Lopt<-candL[i]
              #Uopt<-candU[j]
              #}
              #}
            }
            if(candU[j]==1 & candL[i]!=length(v)){
              pavaUL<-pava(v[2:(candL[i]-1)],decreasing=TRUE)
              pavaAux<-c(pavaLU[length(pavaLU)],pavaUL,pavaLU[-length(pavaLU)])
              if((length(v)-length(pavaAux))!=0){
                #print("##########################################################error")
              }
              mseAux<-sum((v-pavaAux)^2)/length(v)
```

```
              if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
                 pavaUL[1]<=v[candU[j]] & pavaUL[length(pavaUL)]>=v[candL[i]] &
mseAux<mseFin){
                mseFin<-mseAux
                pavaFin<-pavaAux
                Lopt<-candL[i]
                Uopt<-candU[j]
              }
            }
            if(candL[i]==length(v) & candU[j]!=1){
              pavaUL<-pava(v[(candU[j]+1):(length(v)-1)],decreasing=TRUE)
              pavaAux<-c(pavaLU[2:length(pavaLU)],pavaUL,pavaLU[1])
              if((length(v)-length(pavaAux))!=0){
                #print("#########################################################error")
              }
              mseAux<-sum((v-pavaAux)^2)/length(v)
              if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
                 pavaUL[1]<=v[candU[j]] & pavaUL[length(pavaUL)]>=v[candL[i]] &
mseAux<mseFin){
                mseFin<-mseAux
                pavaFin<-pavaAux
                Lopt<-candL[i]
                Uopt<-candU[j]
              }
            }
            if(candL[i]==length(v) & candU[j]==1){
              pavaUL<-pava(v[(candU[j]+1):(candL[i]-1)],decreasing=TRUE)
              pavaAux<-c(pavaLU[2],pavaUL,pavaLU[1])
              if((length(v)-length(pavaAux))!=0){
                #print("#########################################################error")
              }
              mseAux<-sum((v-pavaAux)^2)/length(v)
              if(pavaLU[2]>v[candL[i]] & pavaLU[length(pavaLU)-1]<v[candU[j]] &
mseAux<mseFin){
                mseFin<-mseAux
                pavaFin<-pavaAux
                Lopt<-candL[i]
                Uopt<-candU[j]
              }
            }
          }
        }

      }
    }
  }
  R2<-PV(v,pavaFin)

  return(list(pavaFin,mseFin,candL,candU,Lopt,Uopt, R2))
}

funcionCosinor<-function(datos,time,periodo){#time entre 0y2pi divididos por el periodo

  xx<-cos(time)
  zz<-sin(time)

  fit<-lm((datos)~xx+zz)
  Mest<-fit$coefficients[1]
  bb<-fit$coefficients[2]
  gg<-fit$coefficients[3]
  print(bb)
  print(gg)
  phiEst<-atan2(-gg,bb)%%(2*pi)
  Aest<-sqrt(bb^2+gg^2)

  yP<-(Mest+Aest*cos(time+phiEst))
  R2<-PV(datos,yP )
  rss<- sum (( datos - yP)^2)
  TU<- (-phiEst)%%(2*pi)
  TL<- (pi-phiEst)%%(2*pi)
```

```
  return(list(Mest+Aest*cos(time+phiEst),Mest,Aest,(time+phiEst)%%(2*pi),phiEst, rss,
R2, TU, TL))
}

TSP_Euc_v7<-function(datos,dist_type="Euclidea",datos2p,pesosE=FALSE,pesosMSE=FALSE

,pesosAdd=FALSE,unPeriodo=FALSE,pesos,centrar=FALSE,intentos=25,onlyHeuristica2=FALSE){

  # Isotonic Regression for data matrix
  # estNP<-function1Local_multiple(datos)
  #estNP<-matrix(0,nrow(datos),1)
  # for (i in 1: nrow(datos)){
  #for (j in 1:ncol(datos)){
  #   estNP[i]<- function1Local(datos[i,])

  # }
  estNP<- funcionIntermedia(datos)
  print(estNP[[1]])
  print("salgo de funcion 1")
  estNP_original<-matrix(0,nrow(datos),ncol(datos))
  print("entro en for ")
  for(i in 1:nrow(datos)){
    estNP_original[i,]<-estNP[[i]]#$pavaFin
  }

  if(centrar==TRUE){
    for(i in 1:nrow(datos)){
      datos[i,]<-(datos[i,]-mean(datos[i,]))
    }
  }

  #varianza de genes en 2 periodos
  varianza<-c()
  if(unPeriodo==FALSE){
    for(i in 1:nrow(datos)){
      varianza[i]<-geneVar(datos2p[i,])
    }
  }else{
    for(i in 1:nrow(datos)){

      varianza[i]<-
varianceModelEstimation(datos[i,],estNP_original[i,],length(unique(estNP_original[i,])))
    }
  }

  #matriz de distancias
  mDist<-array(0,dim=c(ncol(datos),ncol(datos),nrow(datos)))
  if(dist_type=="Euclidea"){
    for(g in 1:nrow(datos)){
      for( i in 1:(ncol(datos)-1)){
        for( j in (i+1):ncol(datos)){
          mDist[i,j,g]<-(datos[g,i]-datos[g,j])^2
          mDist[j,i,g]<-mDist[i,j,g]
        }
      }
    }
  }

  if(dist_type=="Coseno"){
    for(g in 1:nrow(datos)){
      for( i in 1:(ncol(datos)-1)){
        for( j in (i+1):ncol(datos)){
          mDist[i,j,g]<-1-cos(datos[g,i]-datos[g,j])
          mDist[j,i,g]<-mDist[i,j,g]
        }
      }
    }
  }

  if(dist_type=="Minkowski"){
    for(g in 1:nrow(datos)){
```

```
      for( i in 1:(ncol(datos)-1)){
        for( j in (i+1):ncol(datos)){
          mDist[i,j,g]<-abs(datos[g,i]-datos[g,j])
          mDist[j,i,g]<-mDist[i,j,g]
        }
      }
    }
  }



  wMSEaux<-rep(1,nrow(datos))
  if(pesosMSE==TRUE){
    for(i in 1:nrow(datos)){
      wMSEaux[i]<-1/varianza[i]
    }
  }

  wMSE<-wMSEaux/sum(wMSEaux)

  wE<-rep(1,nrow(datos))
  if(pesosE==TRUE){
    wE<-wMSE#para incorporar pesos del mse en la matriz de distancias
  }

  wAdd<-rep(1,nrow(datos))
  if(pesosAdd==TRUE){
    for(i in 1:nrow(datos)){
      wAdd[i]<-pesos[i]
    }
    wAdd<-wAdd/sum(wAdd)
  }

  mE<-matrix(0,ncol(datos),ncol(datos))
  if(pesosAdd==TRUE){
    for( i in 1:(ncol(datos)-1)){
      for( j in (i+1):ncol(datos)){
        mE[i,j]<-sum(mDist[i,j,]*wE*wAdd)
        mE[j,i]<-mE[i,j]
      }
    }
  }else{
    for( i in 1:(ncol(datos)-1)){
      for( j in (i+1):ncol(datos)){
        mE[i,j]<-sum(mDist[i,j,]*wE)
        mE[j,i]<-mE[i,j]
      }
    }
  }

  sol<-TSP(as.dist(mE))
  if(onlyHeuristica2==TRUE){
    methods="farthest_insertion"
  }else{
    methods<-c("nearest_insertion", "farthest_insertion", "cheapest_insertion",
"arbitrary_insertion","nn",
               "repetitive_nn",      "2-opt")
  }
  sol1<-list()
  labels_sol1<-list()
  newData<-array(0,dim=c(nrow(datos),ncol(datos),length(methods)))
  estNP_newData<-array(0,dim=c(nrow(datos),ncol(datos),length(methods)))
  sumaMseAfter<-99999999
  suma<-c()

  startRealTime<-proc.time()
  suma3<-c()
  if(pesosAdd==TRUE){
    for ( i in 1:nrow(datos)){
      suma3[i]<-wMSE[i]*wAdd[i]*sum((datos[i,]-estNP_original[i,])^2)/ncol(datos)
    }
```

```
  }else{
    for ( i in 1:nrow(datos)){
      suma3[i]<-wMSE[i]*sum((datos[i,]-estNP_original[i,])^2)/ncol(datos)
    }
  }
  sumaMseOriginal<-sum(suma3)
  realTime<-(proc.time()-startRealTime)[3]

  print(paste("MSE orden real: ",sumaMseOriginal,sep=""))

  for( k in 1:length(methods)){
    for(veces in 1:intentos){
      print(paste("Heuristica ",methods[k],", Intento:",veces,sep=""))
      sol1[[k]]<-solve_TSP(sol,method=methods[k])
      labels_sol1[[k]]<-as.numeric(labels(sol1[[k]]))
      newData[,,k]<-datos[,labels_sol1[[k]]]
      estNP_newData2<- funcionIntermedia(newData[,,k])
      #estNP_newData2<- function1Local_multiple(newData[,,k])
      #estNP_newData2<- function1Local(newData[,,k])
      estNP_newData2_aux<-matrix(0,nrow(datos),ncol(datos))
      for(i in 1:nrow(datos)){
        print(estNP_newData)
        print(estNP_newData2[[i]])
        estNP_newData2_aux[i, ]<-estNP_newData2[[i]] #$pava
      }

      estNP_newData[,,k]<-estNP_newData2_aux
      suma1<-c()
      if(pesosAdd==TRUE){
        for(i in 1:nrow(newData[,,k])){
          suma1[i]<-wMSE[i]*wAdd[i]*sum((newData[i,,k]-
estNP_newData[i,,k])^2)/length(newData[i,,k])
        }
      }else{
        for(i in 1:nrow(newData[,,k])){
          suma1[i]<-wMSE[i]*sum((newData[i,,k]-
estNP_newData[i,,k])^2)/length(newData[i,,k])
        }
      }
      suma[k]<-sum(suma1)

      if(suma[k]<sumaMseAfter ){
        heuristica<-k
        orderFin<-labels_sol1[[k]]
        sumaMseAfter<-suma[k]
        print(paste("Mejorando MSE a partir orden TSP ..... MSEafter:
",sumaMseAfter,sep=""))
        mseBefore<-suma3
        mseAfter<-suma1
      }
    }
  }

return(list(orderFin,heuristica,sumaMseAfter,mDist,mE,suma,sol1,labels_sol1,newData,estN
P_newData,wE,wMSE,mseBefore,mseAfter,sumaMseOriginal,wMSEaux,realTime))
}

funcionIntermedia<- function(datos){
  estNP<-matrix(0,nrow(datos),ncol(datos))
  for (i in 1: nrow(datos)){
    print(i)
    estNP[i]<- function1Local(datos[i,])
  }
  return (estNP)
}

varianceModelEstimation<-function(original,adjModel,nParam){
  return(sum((original-adjModel)^2)/(length(original)-nParam))
}
```

```
cosinor.lm <- function(formula, period = 12,
                       dato, na.action = na.omit){

  # build time tranformations

  Terms <- terms(formula, specials = c("time", "amp.acro"))

  stopifnot(attr(Terms, "specials")$time != 1)
  varnames <- get_varnames(Terms)
  timevar <- varnames[attr(Terms, "specials")$time - 1]

  xx<-cos(dato[,timevar])
  zz<-sin(dato[,timevar])

  fit<-lm((dato[,1])~xx+zz)
  Mest<-fit$coefficients[1]
  bb<-fit$coefficients[2]
  gg<-fit$coefficients[3]
  phiEst<-atan2(-gg,bb)%%(2*pi)
  Aest<-sqrt(bb^2+gg^2)

  rss<- sum (( dato[,1] - (Mest+bb*cos(2*pi*time/periodo)-gg*sin(2*pi*time/periodo)))^2)

  #data$rrr <- cos(2 * pi * dato[,timevar] / period)
  #data$sss <- sin(2 * pi * dato[,timevar] / period)

  data$rrr <- xx
  data$sss <- zz

  spec_dex <- unlist(attr(Terms, "special")$amp.acro) - 1
  mainpart <- c(varnames[c(-spec_dex, - (attr(Terms, "special")$time - 1))], "rrr",
"sss")
  acpart <- paste(sort(rep(varnames[spec_dex], 2)), rep(c("rrr", "sss"),
length(spec_dex)), sep = ":")
  newformula <- as.formula(paste(rownames(attr(Terms, "factors"))[1],
                                 paste(c(mainpart, acpart), collapse = " + "), sep = " ~
"))



  fit <- lm(newformula, data, na.action = na.action)

  mf <- fit

  r.coef <- c(FALSE, as.logical(attr(mf$terms, "factors")["rrr",]))
  s.coef <- c(FALSE, as.logical(attr(mf$terms, "factors")["sss",]))
  mu.coef <- c(TRUE, ! (as.logical(attr(mf$terms, "factors")["sss",]) |
                          as.logical(attr(mf$terms, "factors")["rrr",])))

  beta.s <- mf$coefficients[s.coef]
  beta.r <- mf$coefficients[r.coef]

  groups.r <- c(beta.r["rrr"], beta.r["rrr"] + beta.r[which(names(beta.r) != "rrr")])
  groups.s <- c(beta.s["sss"], beta.s["sss"] + beta.s[which(names(beta.s) != "sss")])

  amp <-Aest
  #amp <- sqrt(groups.r^2 + groups.s^2)
  names(amp) <- gsub("rrr", "amp", names(beta.r))

  #acr <-phiEst
  acr <- atan(groups.s / groups.r)
  # print("acrofase")
  # print(atan(groups.s / groups.r))
  names(acr) <-  gsub("sss", "acr", names(beta.s))
  coef <- c(mf$coefficients[mu.coef], amp, acr)
  #coef <- c(Mest, amp, acr)


  structure(list(fit = fit, Call = match.call(), Terms = Terms, coefficients = coef,
period = period, rss=rss), class = "cosinor.lm")
```

```
}

get_varnames <- function(Terms){

  spec <- names(attr(Terms, "specials"))
  tname <- attr(Terms, "term.labels")

  dex <- unlist(sapply(spec, function(sp){

    attr(Terms, "specials")[[sp]] - 1

  }))

  tname2 <- tname
  for(jj in spec){

    gbl <- grep(paste0(jj, "("), tname2, fixed = TRUE)
    init <- length(gbl) > 0
    if( init ){
      jlack <- gsub(paste0(jj, "("), "", tname2, fixed = TRUE)
      tname2[gbl] <- substr(jlack[gbl], 1, nchar(jlack[gbl]) - 1)
    }

  }

  tname2

}

update_covnames <- function(names){

  covnames <- grep("(amp|acr|Intercept)", names, invert = TRUE, value = TRUE)

  lack <- names
  for(n in covnames){
    lack <- gsub(paste0(n, ":"), paste0("[", n, " = 1]:"), lack)
    lack <- gsub(paste0("^", n, "$"), paste0("[", n, " = 1]"), lack)
  }
  lack
}

ggplot.cosinor.lm <- function(object,l, x_str = NULL){

  timeax <- seq(0, object$period, length.out = l)
  covars <- grep("(rrr|sss)", attr(object$fit$terms, "term.labels"), invert = TRUE,
value = TRUE)

  newdata <- data.frame(time = timeax, rrr = cos(2 * pi * timeax / object$period),
                        sss = sin(2 * pi * timeax / object$period))
  for(j in covars){
    newdata[,j] <- 0
  }
  if(!is.null(x_str)){

    for(d in x_str){

      tdat <- newdata
      tdat[,d] <- 1
      newdata <- rbind(newdata, tdat)

    }
    newdata$levels <- ""
    for(d in x_str){

      newdata$levels <- paste(newdata$levels, paste(d, "=", newdata[,d]))

    }


  }
```

```
  newdata$Y.hat <- predict(object$fit, newdata = newdata)

  if(missing(x_str) || is.null(x_str)){

    ggplot(newdata, aes_string(x = "time", y = "Y.hat")) + geom_line()

  } else {

    ggplot(newdata, aes_string(x = "time", y = "Y.hat", col = "levels")) + geom_line()

  }
}

#datos <- indORIDataReduced
#gen <- "ARNTL"
#longitud <- dim(ori)[1]
#periodo <- 24
#time <- time
#titulo= "grafica"

#datos <- indORIDataReduced1
#gen<- "PER3"
#longitud <- longitud1
#periodo<-24
#time<- time1
#titulo <- "hombre"

grafico<-function(datos, gen, longitud, periodo,time, titulo= "grafica"){

  #\datos <- indORIDataReduced7
  #gen<- "ARNTL"
  #longitud <- longitud7
  #periodo<-24
  #time<- time7
  #titulo <- "hombre"

  ti <- time
  datosCos<-rescale(datos[gen,], to=c(-1,1))
  dataGen<- rescale(datos[gen,], to=c(-1,1))
  NP<- function1Local(dataGen)
  COS<-funcionCosinor(datosCos,time,periodo)
  print(" phi")
  print(COS[[5]])
  FMM<-fitFMM(vData=dataGen,timePoints=time)
  tu<-rescale(rep(1:longitud,1),to=c(0, 2 * pi))
  data<-rbind(datosCos, ti)
  data<-t(data)
  dato<-data
  dato<-as.data.frame(dato)
  cos<-cosinor.lm(dato[,1]~ time(ti),period = periodo, dato=dato )

  g<-ggplot.cosinor.lm(cos, l=longitud)+geom_line(lwd=1,color="red") +
ggtitle(titulo)+theme(aspect.ratio=1)+labs(x = "Tiempo", y =" Expresion del gen")
  g<-g+geom_line(aes(y=NP[[1]]), lwd=1, col=3)#ajuste NP

  print(FMM@M+FMM@A*cos(FMM@beta+2*atan(FMM@omega*tan((time-FMM@alpha)/2))))
  g<-g+geom_line(aes(y=FMM@M+FMM@A*cos(FMM@beta+2*atan(FMM@omega*tan((tu-
FMM@alpha)/2)))), col=4,lwd=1) #FMM
  g<-g+geom_point(aes(y=datosCos))
  #g <-g + annotate("text", x = 18, y = 0.1, label = "Cosinor", colour = "red",size = 3)
  #g <-g + annotate("text", x = 18, y = 0.85, label = "FMM", colour = "blue",size = 3)
  g
  return(g)

}
#(NP[[6]])

#Suaviza el FMM.No funciona siempre, en algunos casos taslada la curva
```

```
graficoSmooth<-function(datos, gen, longitud, periodo,t, titulo= "grafica",
smooth=1000){

  datosCos<-rescale(datos[gen,], to=c(-1,1))
  dataGen<- rescale(datos[gen,], to=c(-1,1))
  NP<- function1Local(dataGen)
  COS<-funcionCosinor(datosCos,t,periodo)
  FMM<-fitFMM(vData=dataGen,timePoints=t)

  data<-rbind(datosCos, t)
  data<-t(data)
  dato<-data
  dato<-as.data.frame(dato)
  cos<-cosinor.lm(dato[,1]~ time(t),period = periodo, dato=dato )
  timeax <- seq(0, 2*pi, length.out = smooth)
  obj2<-generateFMM(FMM@M,FMM@A,FMM@alpha ,FMM@beta,FMM@omega, timePoints = timeax)
  yFMM<-(y=FMM@M+FMM@A*cos(FMM@beta+2*atan(FMM@omega*tan((timeax-FMM@alpha)/2))))
  print(yFMM)
  tFMM<-timeax*(12/pi)
  data2<-data.frame(yFMM,tFMM)
  data2<-data2[data2$yFMM < 1 ,]
  data2<-data2[data2$yFMM > -1 ,]
  print(data2)
  g<-ggplot.cosinor.lm(cos, l=longitud)+geom_line(lwd=1,color="red") +
ggtitle(titulo)+theme(aspect.ratio=1)+labs(x = "Tiempo", y =" Expresi???n del gen")
  g<-g+geom_line(aes(y=NP[[1]]), lwd=1, col=3)#ajuste NP
  g<-g+geom_line(data=data2,aes(x=tFMM, y=yFMM), lwd=1, col=4)#ajuste FMM
  g<-g+geom_point(aes(y=datosCos))
  g
  return(g)

}


# Calcula error R2 parametros y picos
calculoError<- function(data,nGen,nIndv, time, period, NP="NP", Cos="Cos", FMM="FMM"  ){


  error<- matrix(nrow=nGen, ncol = 4)
  ajuste<- matrix(nrow=nGen, ncol = 4)
  parametros_Cosinor<- matrix(nrow=nGen, ncol = 4)
  parametros_FMM<- matrix(nrow=nGen, ncol = 6)
  picos_Cosinor<- matrix(nrow=nGen, ncol = 5)
  picos_FMM<- matrix(nrow=nGen, ncol = 5)

  for (i in 1:nGen){#nrow(indORIDataReducedSCZ)){
    data[i,]<- rescale(data[i,], to=c(-1,1))
    NP<- function1Local(data[i,])
    COS<-funcionCosinor(data[i,],time,period)
    FMM<-fitFMM(vData=data[i,],timePoints=time)
    picoFMM<-getFMMPeaks(FMM)
    a<-ajus(data, i, time, "Ori")

    error[i, ]<- c(rownames(data)[i] ,as.numeric(NP[[2]]),
                   as.numeric(COS[[6]]/nIndv), as.numeric(getSSE(FMM)/nIndv))
    ajuste[i, ]<- c(rownames(data)[i] ,a$NP, a$cos, a$FMM)
    parametros_Cosinor[i,]<-c(rownames(data)[i] ,COS[[2]],COS[[3]], COS[[5]])
    parametros_FMM[i,]<-c(rownames(data)[i] ,FMM@M, FMM@A, FMM@alpha, FMM@beta,
FMM@omega)
    picos_Cosinor[i,]<-c(rownames(data)[i]
,(COS[[2]]+COS[[3]]*cos(pi)),(COS[[2]]+COS[[3]]*cos(0)), COS[[9]], COS[[8]])
    picos_FMM[i,]<-c(rownames(data)[i] ,picoFMM$ZL, picoFMM$ZU,
picoFMM$tpeakL,picoFMM$tpeakU)
  }


  colnames(error)<-c("gen","NpError","CosError", "FMMError")
  colnames(ajuste)<-c("gen","NpR2","CosR2", "FMMR2")
  colnames(parametros_Cosinor)<-c("gen","M","A", "phi")
  colnames(parametros_FMM)<-c("gen","M","A", "alpha", "beta", "omega")
  colnames(picos_Cosinor)<-c("gen","ZL","ZU", "TL", "TU")
```

```
  colnames(picos_FMM)<-c("gen","ZL","ZU", "TL", "TU")

  NPError<-sum(as.numeric(error[,2]),na.rm=TRUE)
  CosError<-sum(as.numeric(error[,3]),na.rm=TRUE)
  FMMError<-sum(as.numeric(error[,4]),na.rm=TRUE)

  return(list(error=error, R2=ajuste, parametros_Cosinor=parametros_Cosinor,
parametros_FMM=parametros_FMM, pico_Cosinor=picos_Cosinor, pico_FMM=picos_FMM))

}


#Calcula R2
PV <- function(vData,pred){
  meanVData <- mean(vData)
  return(1 - sum((vData-pred)^2)/sum((vData-meanVData)^2))
}



ajusteR2<- function (Ori, OriR, ZT, t, tZT, gen){

  ajus(Ori, gen, t, "Ori")
  ajus(OriR, gen, t, "Ori Reducido ")
  ajus(ZT, gen, tZT, "ZT ")

}


ajus<- function(data, gen, time, nombre= "Ajuste"){
  datos<-data[gen,]
  dataGen<- rescale(datos, to=c(-1,1))
  NP3<- function1Local(dataGen)
  FMM<-fitFMM(vData=dataGen,timePoints=time)
  cos<- funcionCosinor(dataGen, time, 24)
  return(list(NP=NP3[[7]], cos=cos[[7]], FMM=FMM@R2))

}


escalado<- function(datos){
  esc<- matrix(nrow=1, ncol = length(datos))
  for( i in 1: length(datos)){

    esc[i]<- (datos[i] + 6)/(12/pi)

  }
  esc<-c(esc)
  return(esc)

}


escaladoInverso<- function(datos){
  esc<- matrix(nrow=1, ncol = length(datos))
  for( i in 1: length(datos)){

    esc[i]<- ((datos[i] *12)/pi)

  }
  esc<-c(esc)
  return(esc)

}

generateFMM <-
function(M,A,alpha,beta,omega,from=0,to=2*pi,length.out=100,timePoints=seq(from,to,lengt
h=length.out),
                         plot=TRUE,outvalues=TRUE,sigmaNoise=0){

  pl<-nullGrob()
```

```
  narg <- max(length(M),length(A),length(alpha),length(beta),length(omega))

  if(length(M)>1){
    warning("M parameter should be a vector of length 1.
            The intercept parameter used in the simulation is the sum of the elements of
the argument M.")
    M <- sum(M)
  }
  M <- rep(M/narg,length.out=narg)


  A <- rep(A,length.out=narg)
  if(sum(A <= 0) > 0) stop("A parameter must be positive.")

  alpha <- rep(alpha,length.out=narg)
  alpha <- alpha%%(2*pi) # between 0 and 2*pi

  beta <- rep(beta,length.out=narg)
  beta <- beta%%(2*pi) # between 0 and 2*pi

  omega <- rep(omega,length.out=narg)
  if(sum(omega<0)>0 | sum(omega>1)>0) stop("omega parameter must be between 0 and 1.")

  t <- timePoints

  phi <- list()
  for(i in 1:narg){
    phi[[i]] <- beta[i]+2*atan(omega[i]*tan((t-alpha[i])/2))
  }

  ym <- list()
  for(i in 1:narg){
    ym[[i]] <- M[i]+A[i]*cos(phi[[i]])
  }

  y <- rep(0,length(t))
  for(i in 1:narg){
    y <- y + ym[[i]]
  }

  if (sigmaNoise > 0) y <- y + rnorm(length.out,0,sigmaNoise)

  if(plot) {
    type_<-ifelse(sigmaNoise==0,"l","p")
    pl<- plot(t,y,type=type_,lwd=2,col=2,xlab="Time",ylab="Response",
              main=paste("Simulated data from FMM model"))

  }


  if(outvalues) return(list(input=list(M = M[1]*narg,
A=A,alpha=alpha,beta=beta,omega=omega,t=t,y=y, p=pl)))
}
```