Data Handling and Visualization

Final Visualization

Cristina Baron Suarez
Student ID: 23069038
cb24ack@herts.ac.uk

January 5, 2025

Context

I collaborate with a research team at the University of Malaga. This group has been following older adults annually using accelerometry. They use accelerometers for approximately ten days, and it is during this time that the activities of the patients are measured. The group wants to identify patterns that extend beyond this period, but the number of accelerometers is limited. Therefore, several patients share the same device.

The main goal of this project is to know whether it is possible to detect signals of different frequencies in the combined accelerometry data for the observed period, even if the patients have been swapping accelerometers among themselves.

The study has data from 2014 to 2024. However, in 2015 there was a lack of funding, so data was collected only for a few months, and in 2020 COVID emerged and altered data collection. Therefore, we will limit our analysis to the accelerometry data collected between both periods.

Exploring and transforming the data

	code	followUpYear	file	date	bedTimeStarts	bedTimeMiddle	bedTimeEnds	bedDuration	bedSIBDuration	dur_walking	ENMO_walking	steps_walking
0	P201001	2.0	01001_02__017698_2016-11-29_13-25-12	2016-11-22	-1.20	3.60	8.40	9.60	6.40	175.0	62.86	8647.2
1	P201001	2.0	01001_02__017698_2016-11-29_13-25-12	2016-11-23	0.63	4.01	7.38	6.75	4.13	217.0	58.26	9540.0
2	P201001	2.0	01001_02__017698_2016-11-29_13-25-12	2016-11-24	0.08	3.91	7.73	7.65	4.83	317.0	111.95	22403.0
3	P201001	2.0	01001_02__017698_2016-11-29_13-25-12	2016-11-25	-1.00	3.21	7.42	8.42	4.67	253.0	90.47	15567.0
4	P201001	2.0	01001_02__017698_2016-11-29_13-25-12	2016-11-26	-0.07	4.60	9.27	9.33	5.25	189.0	62.95	6995.0

For each patient, we have several fields:

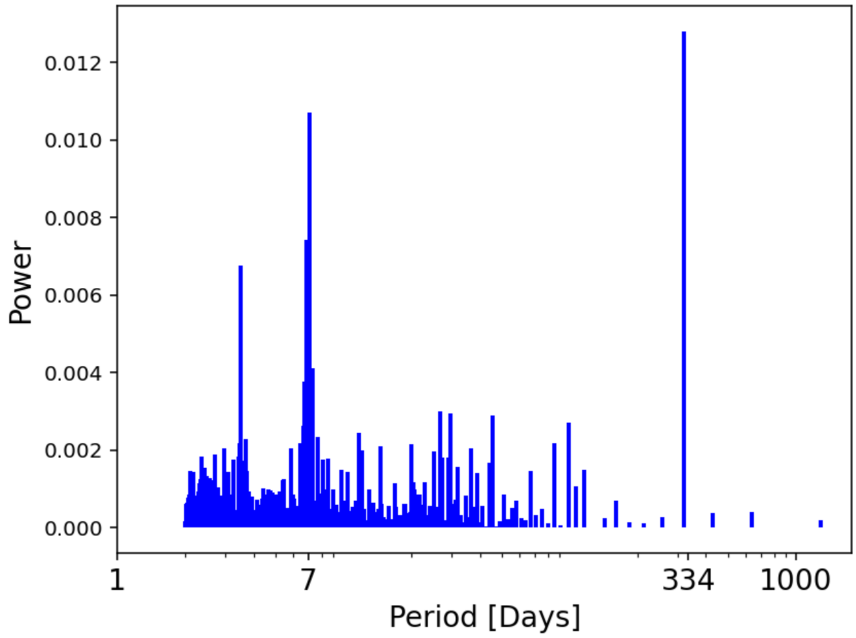
- "code": identifies the patient
- "date": indicates the day the measures were taken
- Set of variables automatically calculated from the day's summary, such as the time they got out of bed, steps taken,...

Let us set the variable 'bedTimeEnds' as a unique value for each day of the studied period. For that, we are going to use the mean collected from all the accelerometers that were active during the selected period of time. As the patients are different, and each of them can have a specific hour of waking up, that is going to introduce significant variability. To reduce it, we will consider how far each individual **deviates** from their own personal mean by subtracting the average observed over the entire study in each participant from the value observed on a specific day.

	date	bedTimeEnds	deviation
0	2016-03-15	9.090000	0.114759
1	2016-03-16	8.689333	-0.051349
2	2016-03-17	8.303913	-0.354033
3	2016-03-18	8.595217	-0.062729
4	2016-03-19	8.694545	0.161345

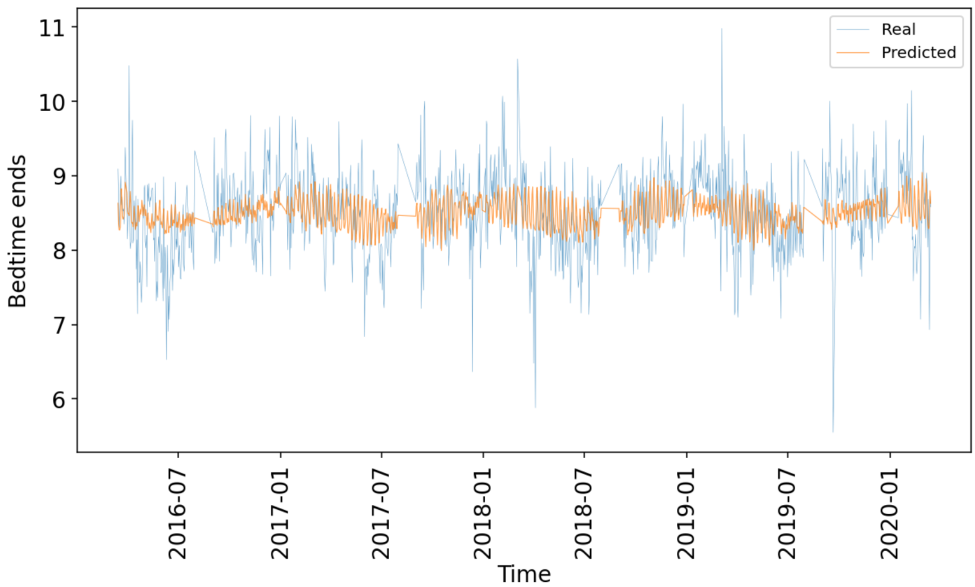
Fourier transform of the variation in people's sleeping schedules over time

We are now going to calculate the power spectrum showing the contribution of different periods (from one week to one year) to the daily variation in awakening.



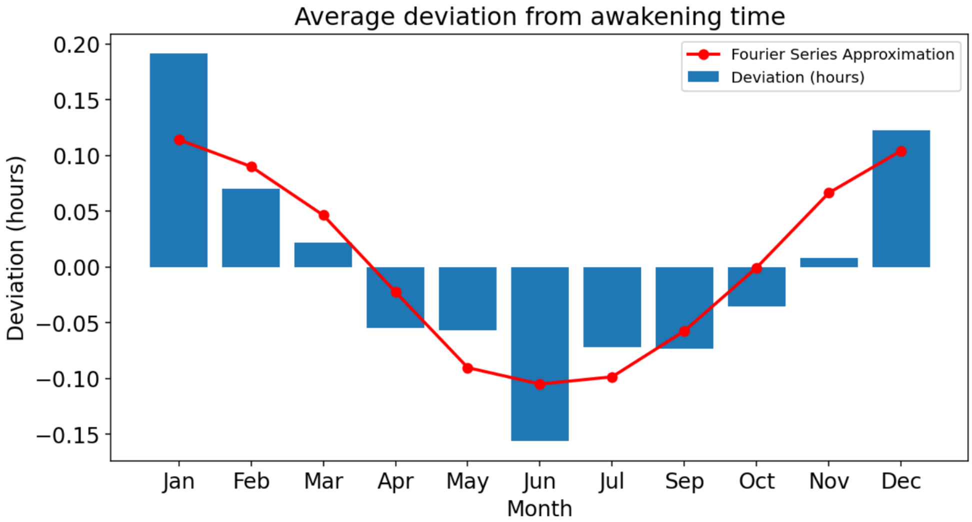
It seems that the 'deviation' is related to periodic components of 320 days (**we have removed 31 days from August**- since no data is collected during this month-, so almost a year) and 7 days. The reason why the longest period can be observed over the course of the year could be probably explained by the time of sunrise. Furthermore, the seven-day cycle is related to the fact that people tend to wake up differently on weekends compared to Monday through Friday.

Having identified the most significant frequencies, and they seem to make sense. Let us visualize how the selected frequencies model the deviations.

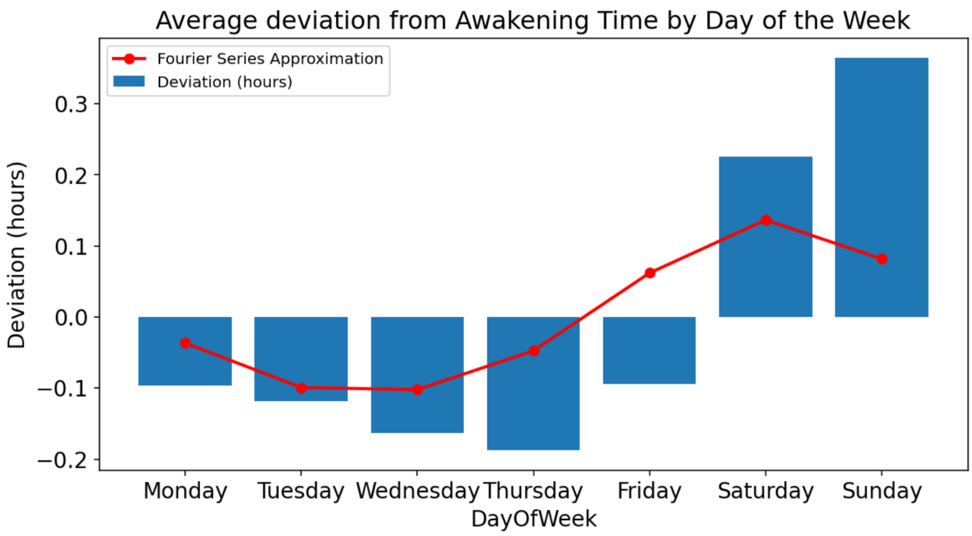


From the previous plot, we can appreciate the yearly and weekly patterns in the dataset, but of course we have chosen only the first most important frequencies.

Now we show the average result for each month, except for August.



Apparently, the deviations in awakening time are related to the changes in daylight throughout the year, although in June the elderly wake up much earlier than we would expect by the data seen in neighboring months and the Fourier model. This may be explained because in Spain the school timetables change, but not the parents' working schedules, and it is a month in which many families need the help of grandparents to take the children to school.



The observed weekly pattern is expected: People wake up later on weekends.

On the other hand, the Fourier approximation struggles to estimate properly the transition to weekends, as is too abrupt to be discerned with the limited number of frequencies that we have chosen.