Ongoing research's report Part I - The core

Project
The main target of this project is to search for predictive algorithms of wrist actigraphy data, for unrestricted lifestyle.

That means that the data analysis is not focusing on the description of the recorded data, but on the search inside them for parameters that could predict short or long term performance.
Those parameters should also be extracted from recordings in real life, not in laboratory settings, in order to use them in telemonitoring services.

For the reader without experience of actigraphy recordings, please read page 11 of this report for a simple model and refer to
http://www.aasmnet.org/Resources/PracticeParameters/PP_Actigraphy_Update.pdf and
https://www.researchgate.net/publication/49758054_The_role_and_validity_of_actigraphy_in_sleep_medicine_An_update for clinical applications.

The first step has been to look for suitable recordings. Most published articles use recordings from few days to few weeks. The reason behind the selection of the length of the recording seems functional to the single research structure without general guidelines on methodological issues for actigraphy. Longer recordings, those needed to develop algorithms, seems rare.
Therefore, late 2015, I set to start a long term recording on myself using a MotionWatch8 recorder (CamNtech Ltd) on the nondominant wrist. All numerical examples in this report are based on those data. Later, an open contest on that recording has been proposed on www.medicoimpianti.it, where data are freely available.

How to analyse data?
The actigraphy recording is usually studied as a time series, for example as in Table I.

Table I

Time		Activity
Day	Hour, Minute	value (counts)
1	00,00	xx
1	00,01	0
1	00,02	0
1	00,03	0
1	00,04	zz
1	00,05	ww
1	00,06	0
1	00,07	0
1	00,08	0
…..		
1	0,m	0
1	0,m+1	0

Inside any time series, there is a positional information that is somehow included in the time value and it is implicit in the time series data analysis.

I suggest to extend Table I, let the positional information be explicit as well as the sequence of zero and non zero series, as in Table II.

Table II

Column A	Column B	Column C	Column D	Column E	Column F
Samples sequence	Time sequence Day Hour, Minute	Activity value (counts)	Series	Series sequence	Length
1	1 00,00	xx	1 of a series of 1	1	1
2	1 00,01	0	1 of a series of 3	2	3
3	1 00,02	0	2 of a series of 3		
4	1 00,03	0	3 of a series of 3		
5	1 00,04	ww	1 of a series of 2	3	2
6	1 00,05	zz	2 of a series of 2		
7	1 00,06	0	1 of a series of m-5	4	m-5
8	1 00,07	0	2 of a series of m-5		
9	1 00,08				
…….	…….				
m+1	1 00,m	0	m-5 of a series of m-5		
m+2	1 00,m+1	yy		5	1

Columns B and C provide the standard time series of Table I shown in Fig. 1 (Y axis =Movement in Counts, X axis= watch time in minutes).

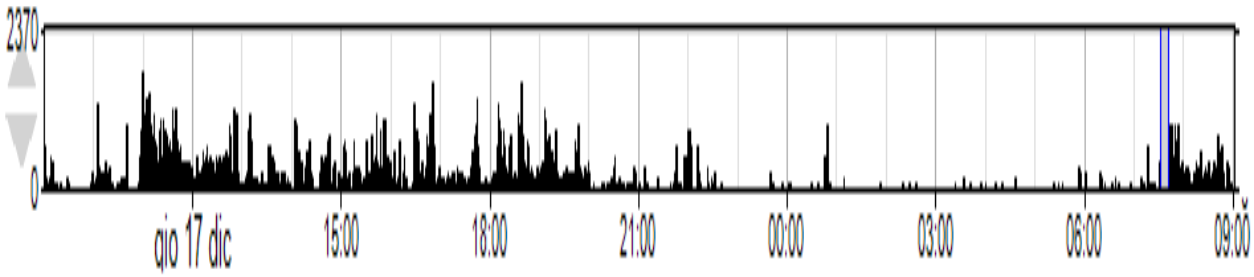


Fig.1 Day 1 Movement counts , 1440 minutes

Columns E and F provide a new graph of 267 series as in Fig. 2 (Y axis =Length in minutes, X axis= Sequence of the series). Note: by definition zero and non-zero series must alternate.

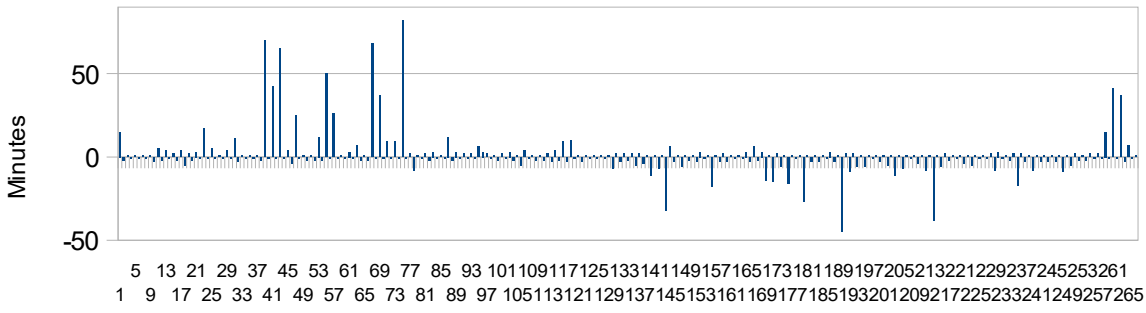


Fig.2 Day 1, Sequence of Non zero (positive) and Zero(negative) series

From Table II, it is possible to relate the two graphs, for instance series n.152 starts at midnight. The counts raw data are the same, but some information is easier to see.
We may use a musical analogy: if we compare the actigraphy recording with a music score, until now pauses have been discarded as meaningless instead of being a rest beat, part of the dynamics. Fig. 2 shows better the different performance of day and night: a different “rhythm” is being played.

As far as I know, (January 2016) no evaluation of those sequences has been made in the past and it is a new field of studies.

Part of this project developed on wrist actigraphy data, shares with HRV (Heart rate variability) the same fundamental statement: “the distance of two involuntary muscles actions is variable, but not random”.
HRV studies the electrical activity related to the heart contractions, actigraphy studies the wrist movement related to skeletal muscles contractions, both mostly involuntary.
Like for HRV 40 years ago, faster sampling and longer recordings allow actigraphy today to evaluate new information.

<u>Ongoing research's report Part II</u> - Pen and pencil example of data analysis

The full analysis of the relationships inside Table II is suitable for AI methodologies.

Some very rough paper and pencil evaluations are however possible.
If we look at zero and non zero values (counts) in the first day of the recording, we find, over the 1440 minutes of the day, 884 (61%) minutes with values and 555 with zeros, grouped in 267 series.

We may start from a simple description and ask ourselves how many series of each length there are.

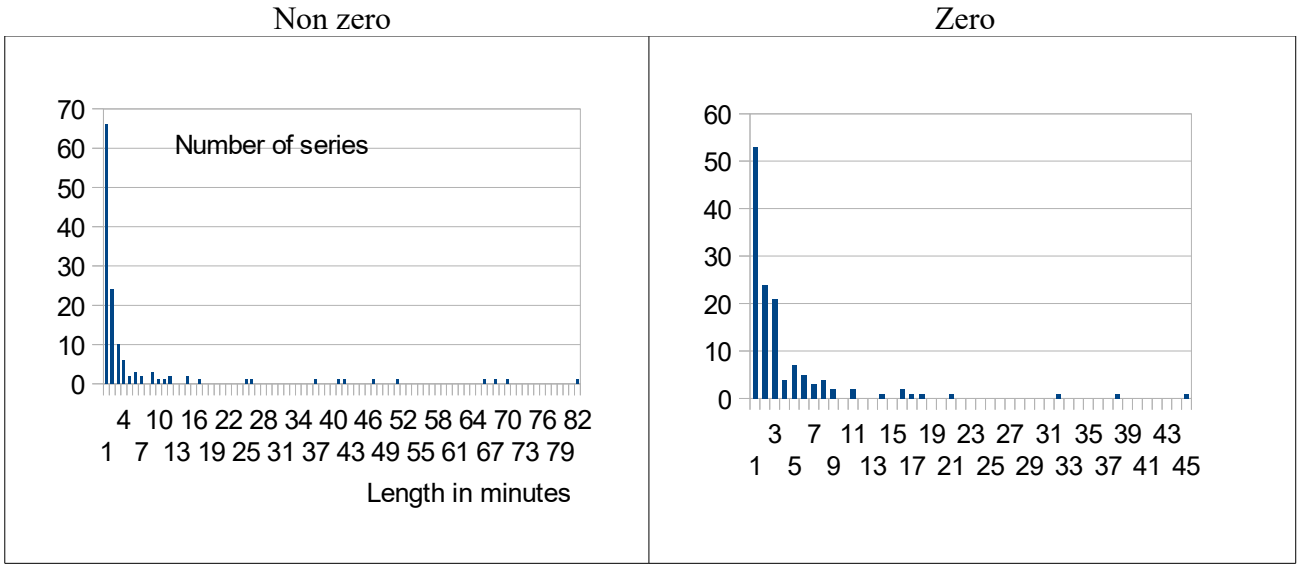


Fig. 3 Day 1 – Number of series in each length.

And compute how much time is spent in each length.

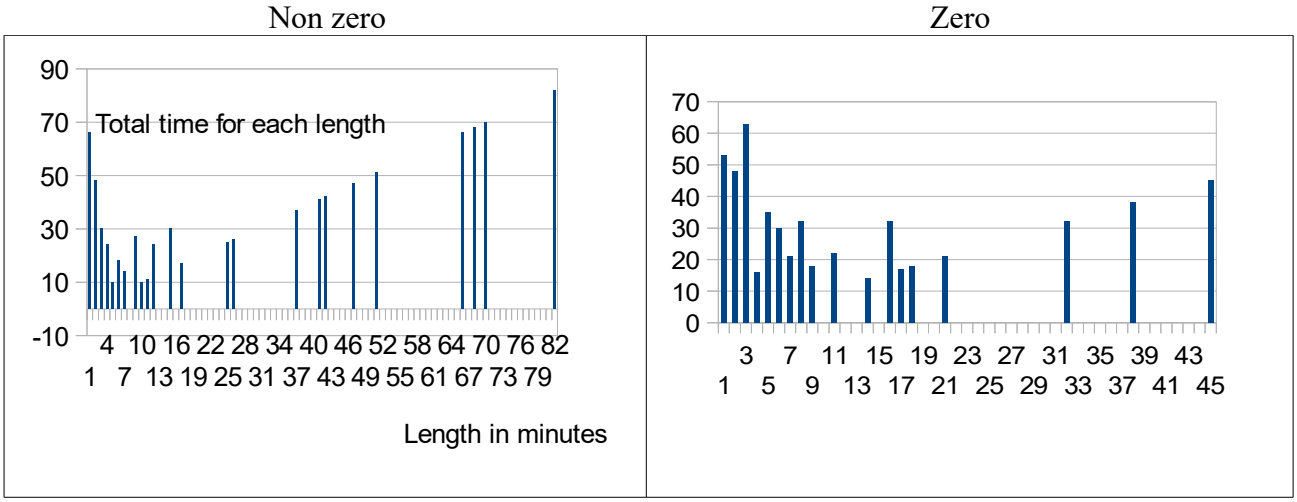


Fig. 4 Day 1 – Total time spent in each length.

For the non zero series, 329 minutes (37%) out of 884, are inside short lengths from 1 to 17 minutes long. The others are concentrated in few, longer series up to 82 minutes each (Fig.4 left column).
Zero minutes (555) are mostly (440, 80%) grouped in short series, up to 21 minutes each, and then few longer ones (Fig.4 right column).

Distributions of that total time as in Fig.5.

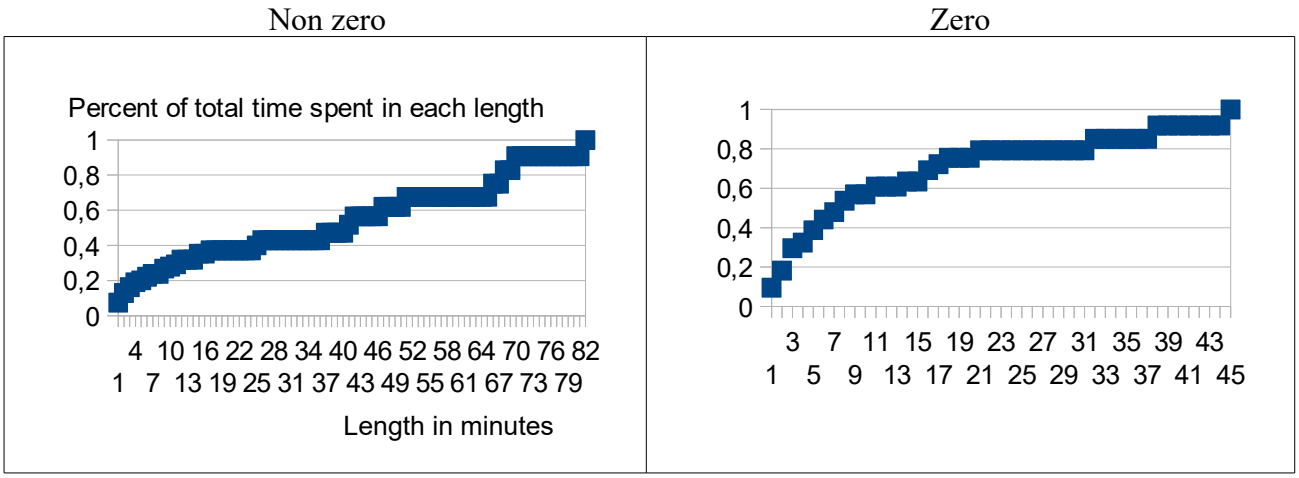


Fig. 5 Day 1 – Total time percent spent in each length.

Are those distributions stable over time? It seems so, at least for the first month.

For the non-zero series, we may describe, for instance, the percent of the counts for each length as in Fig. 6 (Y axis= percent of the counts of the day, X axis=length of the series in minutes).

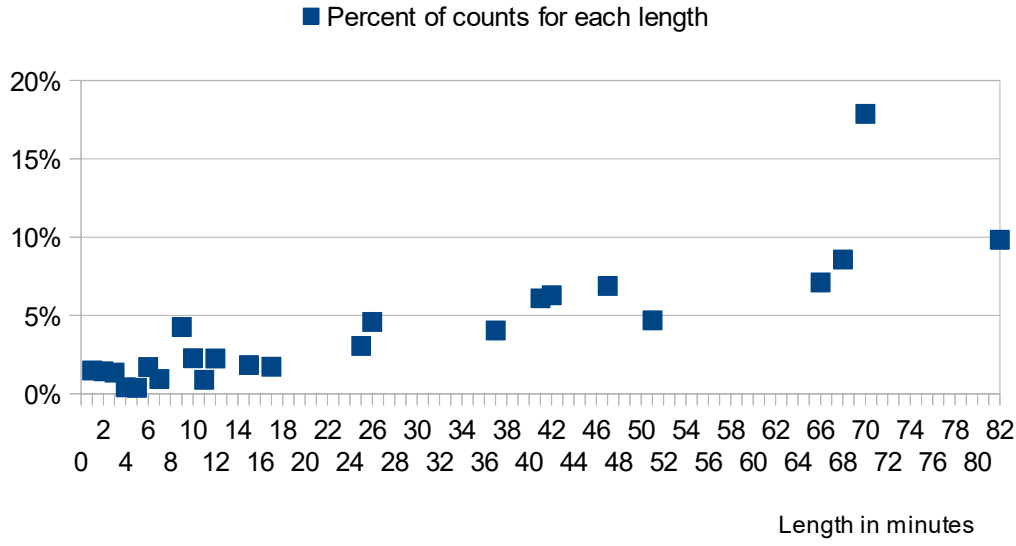


Fig. 6 Day 1 – Percent of counts of the day for each length.

Or the mean value of one minute for each length.

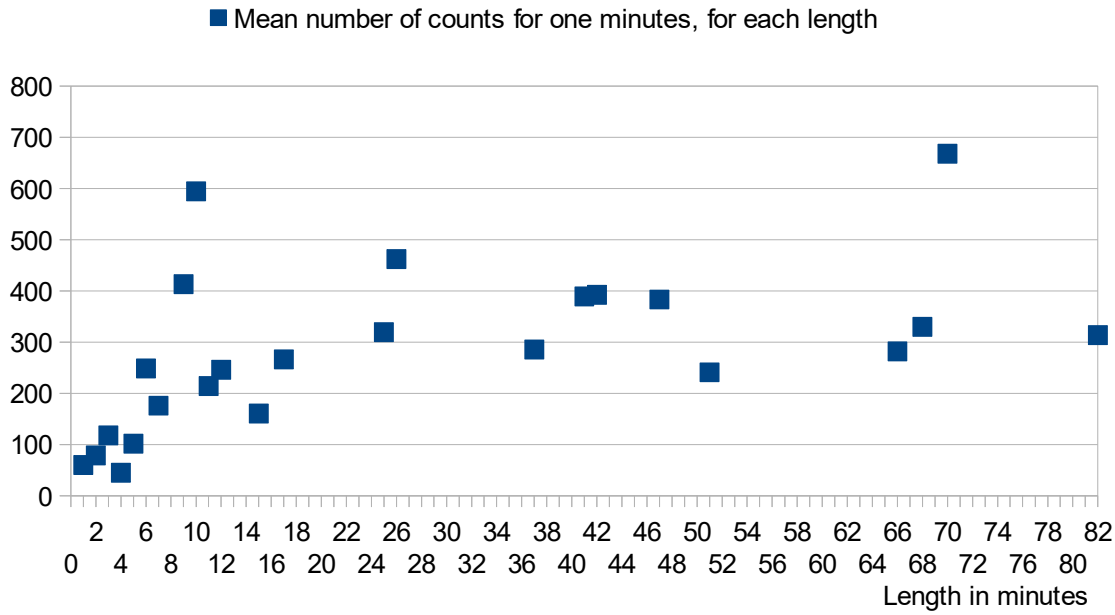


Fig. 7 Day 1 – Minute mean counts for each length.

But, in order to analyse those data we need to extract some feature to compute.
Pen and pencil computation means that we have to focus on few features.
For the selection of them, we need research hypothesis.

First research hypothesis:
"There is something (for the body) tied to the length of the zero series and series with contiguous lengths share that something. "

Segmentation of data brings some errors more or less important, depending on the parameters you are interested in.
If segmentation is needed or useful, in a physiological system we should use an adaptive one.
In this simple example, data around noon are expected to be the ones more influenced by social interactions and to destroy there the continuity is hopefully not to damage too much.
In the worst case probably we would break a long non-zero series.
From that, I segment the samples (Column A of Table II in the first part of the report) in 24 h pieces, starting from noon.

Not every length is present each day, but we need a number to compute each day. For that purpose, I start from the less frequent zero series length, the longest, and search until I find a level that will provide at least one value each night.
On the first month of my recording (27 days, one minute epoch), I find that level at 30 minutes, i.e. each night there is a zero-series long at least 30 minutes. Let's call that group of lengths >=30 minutes, D1.
Using that grid, we get a number each night, but more than one series in the D1 group could be found in a night, up to four in that month, as in Fig.8.

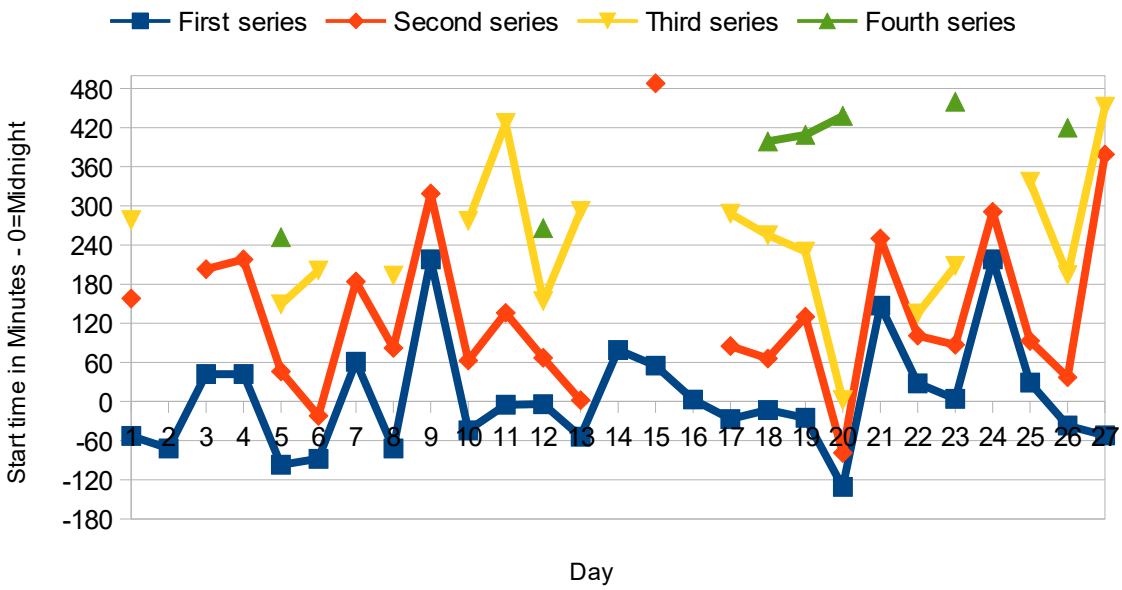


Fig.8 Month 1 D1 components vs Start time

For instance, at night 9 there are 2 D1 series. The first starts at 03.38 and the second at 05.19 .

If we plot all D1 segments of the first month (Fig.9), it seems that extra long segments (> 60 minutes) are allowed to start only inside a time interval.

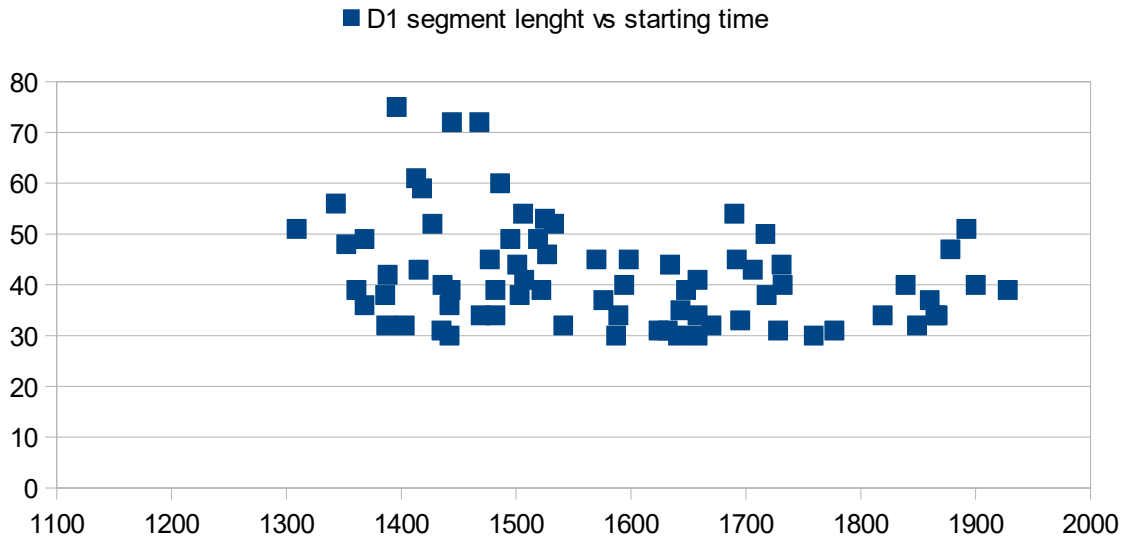


Fig. 9 Month 1 x axis 1440 = midnight

Stretching the first research hypothesis, we may say that the time spent in D1 series in one day is functionally similar up to the point to consider it as a single segment and sum the time as in Fig.10.

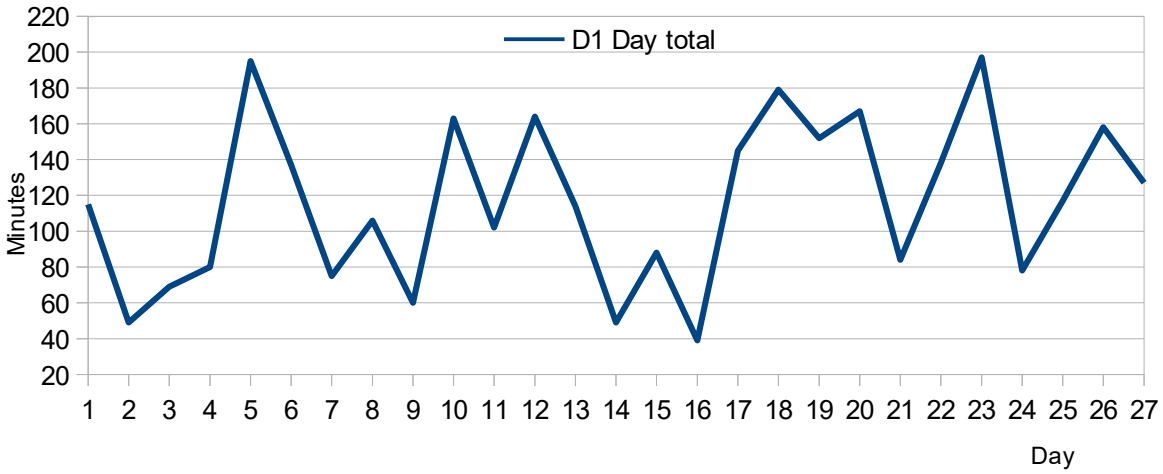


Fig. 10 Month 1, D1 day totals

We may also tie the D1 day total time to the starting time of the first segment and plot the first month as in Fig.11

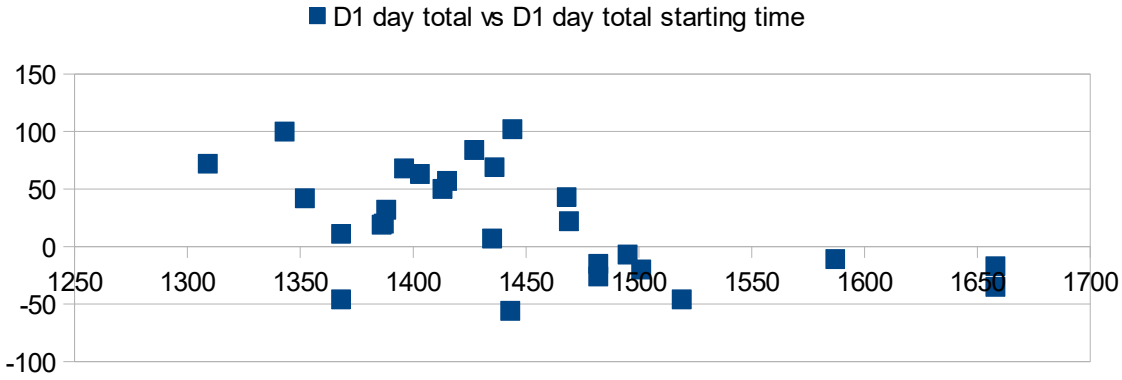


Fig. 11 Month 1 - x axis 1440 = midnight; y axis 0 = mean D1 total time of the month

It seems that, if the D1 first segment is starting after a well defined time, the total value of D1 will stay below the mean of the month.
If we imagine D1 as an estimation of a real sleep characteristic, there is a practical consequence: if we go to bed after that time limit, we are sure that we will not get the minimum D1 related sleep for that night.

After features extraction, we need a data model in order to analyse them.

Second research hypothesis
"We may start from a data model of D1 with two components: one fixed and one variable".
We may use the mean of D1 day total value over an interval as estimation of the fixed part and compute a day-by-day difference from D1 day total value.
If we sum those differences for the first month, we find Fig.12

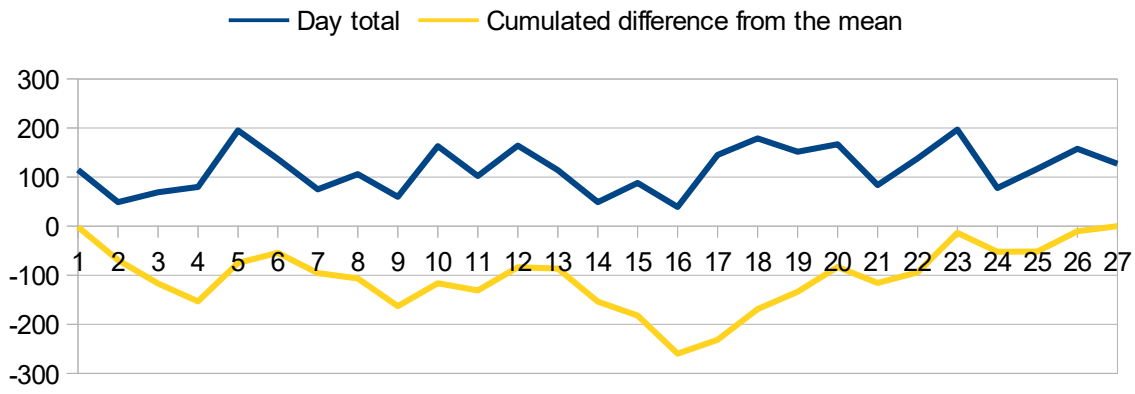


Fig. 12 Month 1 D1 mean=116,5

It is possible to note that the longest D1 day total time are tied to higher accumulated distances from the mean. See day 5, 10, 17. It seems a "reaction", but last days (see days 23 and 26) do not seem coherent with that hypothesis.
Note. By chance the first D1 day total value was near the mean of the month

D2 - D3

If we apply again the grid of the first research hypothesis (at least one value each night), below 30 minutes, we get two more groups: from 16 to 29 minutes and from 9 to 15, that I'll call D2 and D3.

Some D2 segments may appear earlier and later than D1 segments (compare Fig.13 and 9) and D3 even more so (Fig.14).

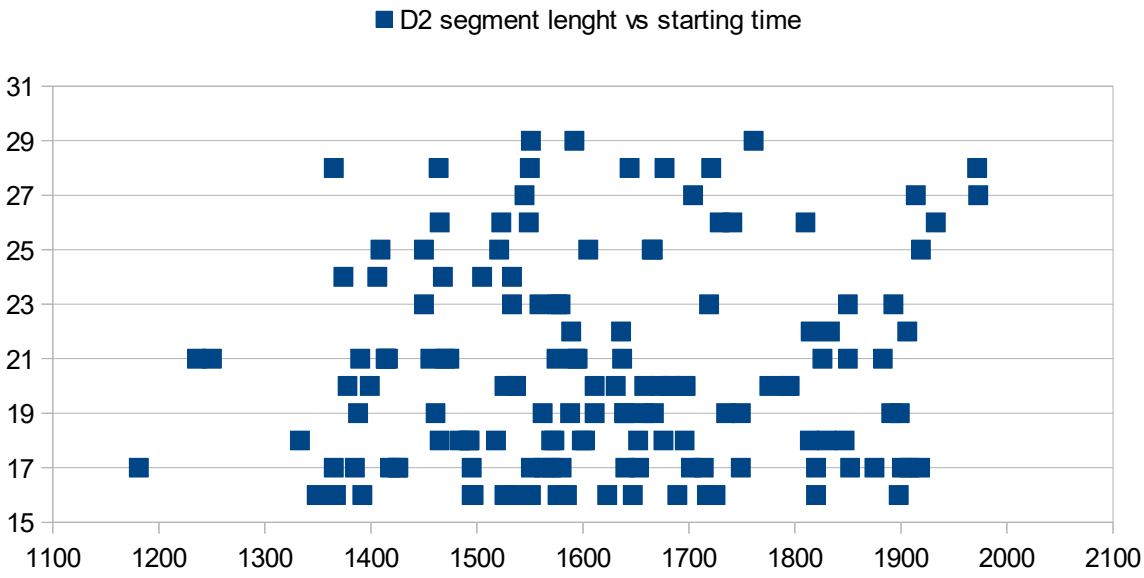


Fig. 13 Month 1 D2 , x axis 1440 = midnight

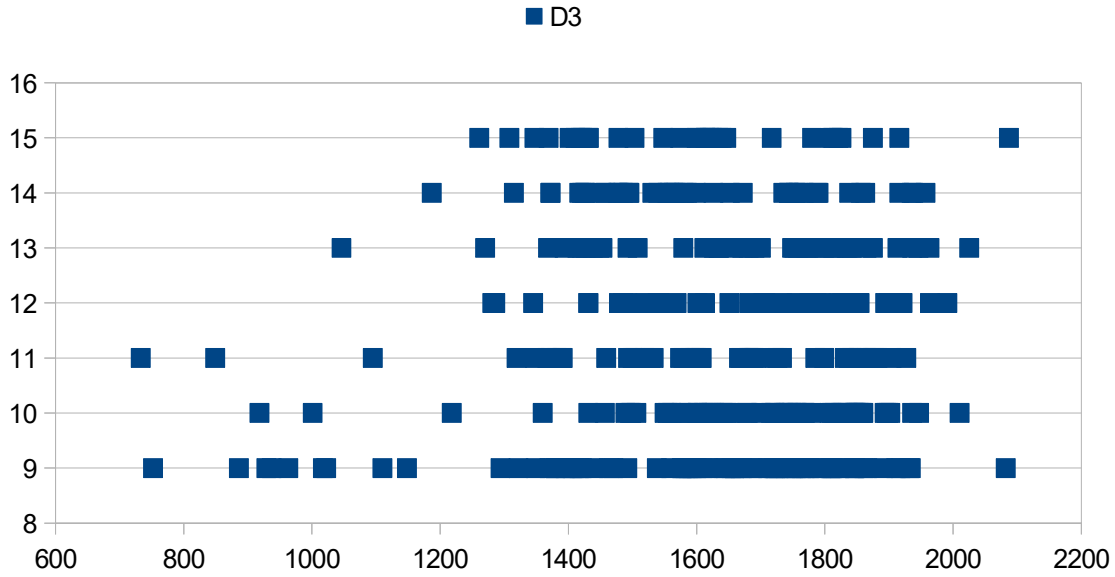


Fig. 14 Month 1 D3 , x axis 1440 = midnight

Terms

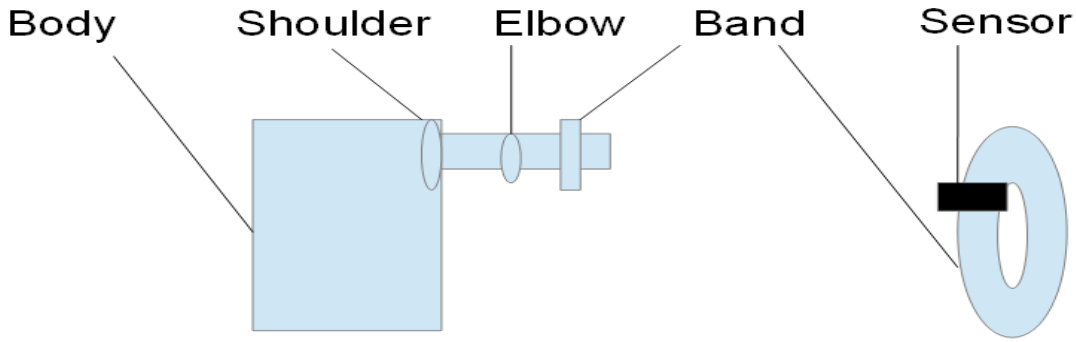
Unfortunately, there are many unknown interferences on the intensity of the activity values.
Few words about “active” and “not active” definitions are needed.

Literature is often confusing with statements like “...wrist actigraphy detects body movements....” or “...we compute activity of the body...”.
It is easy to believe that “active” means “voluntary body moving” and that faster movements will give higher values.
It is not so for the wrist actigraphic data.

The trunk of the body **B** (Fig.15) moves for internal reasons **Dint** (brain status as sleep, body status as muscles position,...) and for environment interactions **Dext** (work, danger, transports, ...).

The sensor is positioned on the nondominant wrist and the limb has is own movement relative to the trunk, **Drel.**

Moreover, the position of the sensor changes around the wrist in an unknown way **M**, as any wristwatch. Therefore, the spatial axis direction xt,yt,zt of the transducer are a time dependent function of the x,y,z of the body.



$$B(t,x,y,z) = Dint(t,x,y,z) + Dext(t,x,y,z)$$

$$S(t,x,y,z) = B(t,x,y,z) + Drel(t,x,y,z) + M(t,x,y,z)$$

Fig.15 Sensor position

When we note a change in the position **S** of the recorder , we do not know “what” (Dint or Dext) was moving and the relationship between S and B.

Moreover, the recorder (Motionwatch8 in my case) samples and filters the signal (S acceleration) in his own way and that also impacts the results (counts).

For the above reasons, in this report I prefer to use the words “zero” and “non zero” instead of “active” and “not active”.

Ongoing research's report Part III - Comments and notes

After 3 months

Note 1) If we extend the analysis of Fig. 11 to the first 3 months (101 values), we find Fig.16.

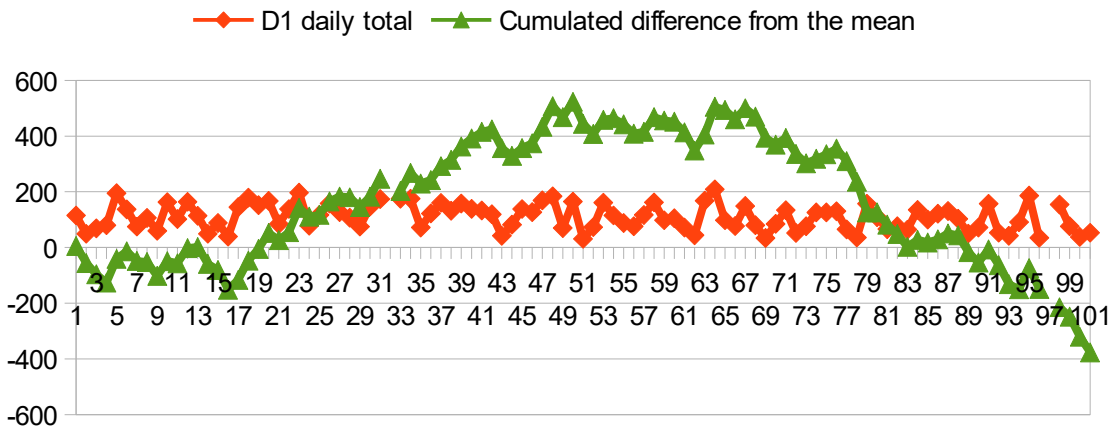


Fig. 16 Month 1 to 3, D1 mean=109,8

It seems that, while the D1 daily total values are always fluctuating inside the 30-200 range, there is a strong shift of the mean expected value.
That would mean that the "fixed" component of our data model will not be a fixed value, and the D1 value of day 23 in Fig. 11 has now a possible new explanation: the high value is not due to a peak of the distance from the mean, but to the increase of the slope of a long trend shift of the mean.

Note: The computed parameters of Fig.16 are limited by the fact that 2 nights do not record D1, days 32 and 97. It seems that the 30 minutes lowest value found in the first month is semi-casual.

After 6 months (end of Project Phase I)

Note 1) "Is there any physiology in that information space?". To answer that question during the night, there is the need to compare with the sleep studies gold standard: polysomnography. Polysomnography uses several signals scored on "pages" of 30 seconds, while actigraphy "history" is made of one minute "epochs".

More then two decades of studies relate wrist actigraphy to sleep and state a strong connection in different situations, using raw data that link one number from actigraphy to two pages of polysomnography.

There is a gap both of sampling and scoring for actigraphy and polysomnography that today has no more reasons to exist.

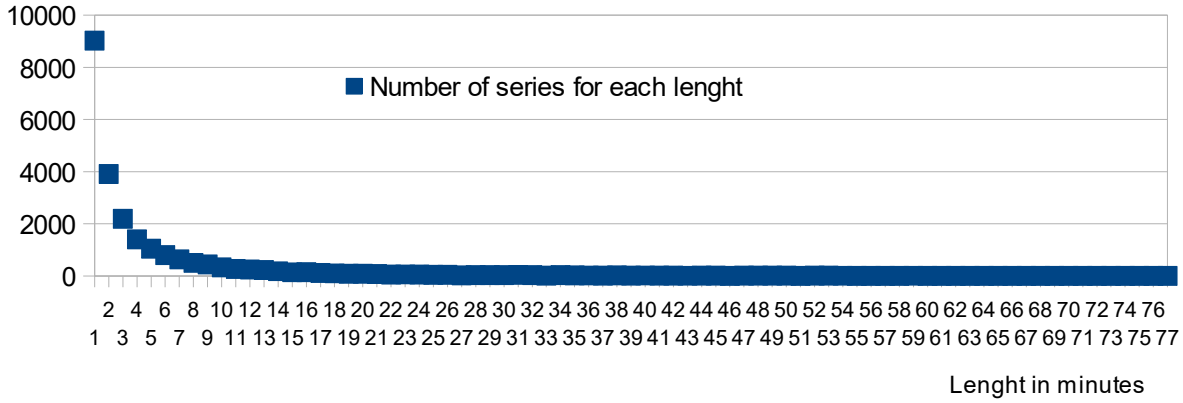
Therefore, from the second semester I set my recording to 1 second "epoch".
That would allow a nice synchronisation of the two methods.

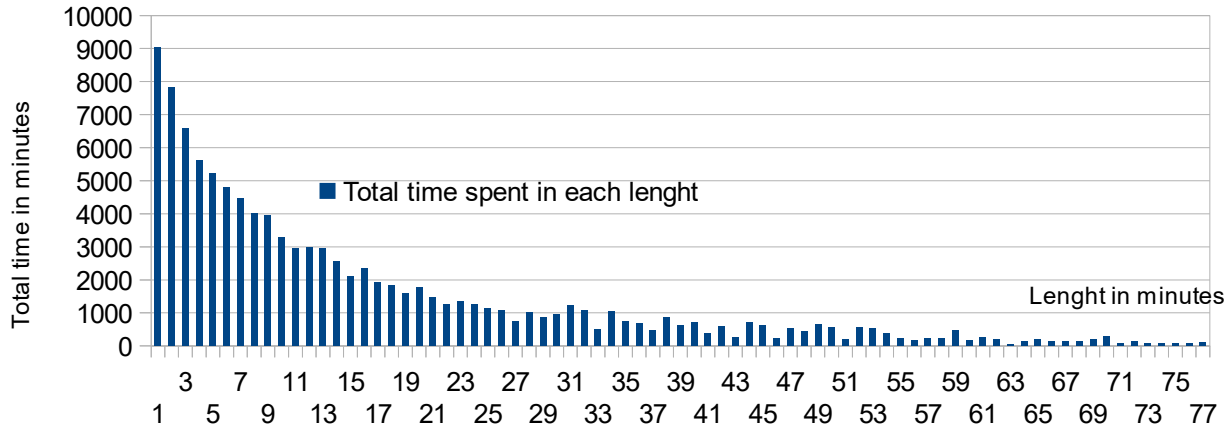
Table II becomes Table III, changing column B to seconds.

Table III

Column A	Column B		Column C	Column D	Column E	Column F
Samples sequence	Time sequence Day	Hour, Minute,Second	Activity value	Series	Series sequence	Length
1	1	00,00,00	xx	1 of a series of 1	1	1
2	1	00,00,01	0	1 of a series of 3	2	3
3	1	00,00,02	0	2 of a series of 3		
4	1	00,00,03	0	3 of a series of 3		
5	1	00,00,04	ww	1 of a series of 2	3	2
6	1	00,00,05	zz	2 of a series of 2		
7	1	00,00,06	0	1 of a series of m-5	4	m-5
8	1	00,00,07	0	2 of a series of m-5		
9	1	00,00,08				
…..						
m+1	1	00,00,00,m	0	m-5 of a series of m-5		
m+2	1	00,00,m+1			5	…..

Note 2) If we perform over the 6 months the evaluation of Fig.3,4, and 5 for the zero series, we find an unexpected very regular distribution with all lengths present.





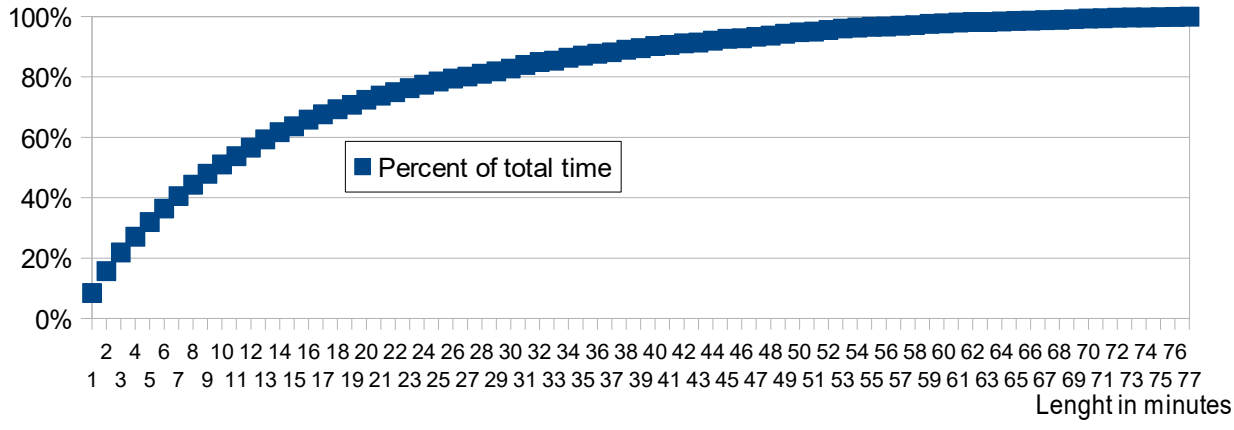


Fig. 17,18 19 Month 1 to 6, zero series

Zero series with length up to 8 minutes form the 44,3 % of the total time, D3 (9-15) is 18,1% , D2 (16-29) is 18,2%, D1 (=>30) is 19,4%.

Note 3) Out of 259.200 minutes of the 6 months, 107.354 (41,4%) are recorded as 0. More or less the same percent of day 1. How much is that percent stable? Over the first month, the minimum is 37% and the maximum is 55%, with a mean of 43,3% as shown in Fig.20.

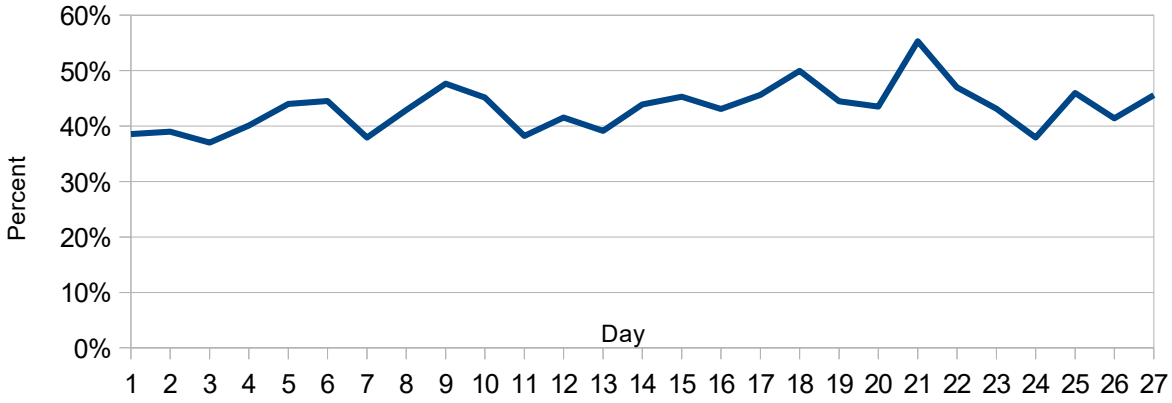


Fig.20 Month 1 – Percent of minutes with zero counts in a day, one minute epoch.

After 9 months

Note 1) Looking at 1 second epoch recordings, it seems that the quantity of the signal in one minute epoch is due to more non zero seconds, not to more strong movement, at least in this lifestyle.

Moreover, movement is seldom long one minute, both day and night as in the example of Fig.21, day on top, night below

.

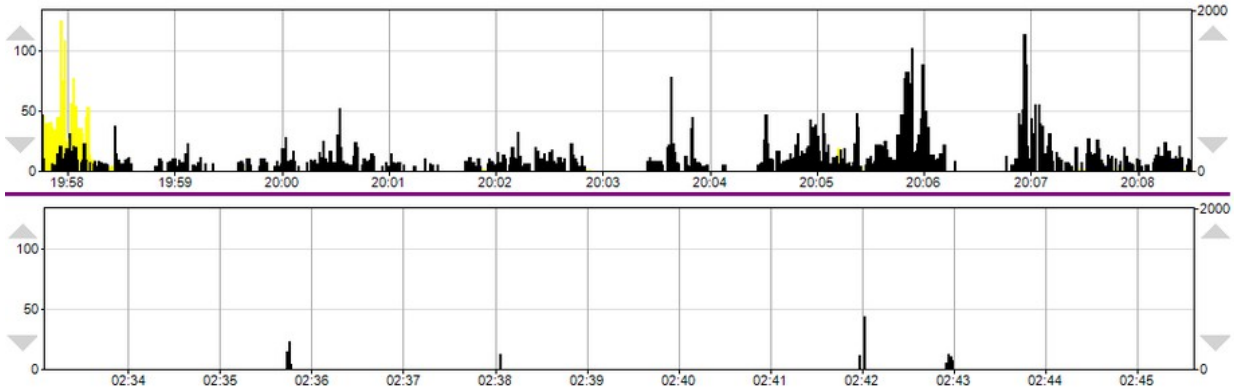


Fig.21 Data with one second epoch. Grid one minute.

That would mean that all studies that are trying to relate activity during the day to sleep at night from wrist actigraphy using 1 minute epoch, over estimate activity time, especially during the night.

After 12 months (end of Project Phase II)

Note 1) All 812 D1 series of the last 12 months are always inside bedtime sleep (Fig.22). If so, we have a strong connection between D1 and sleep that worth investigating.

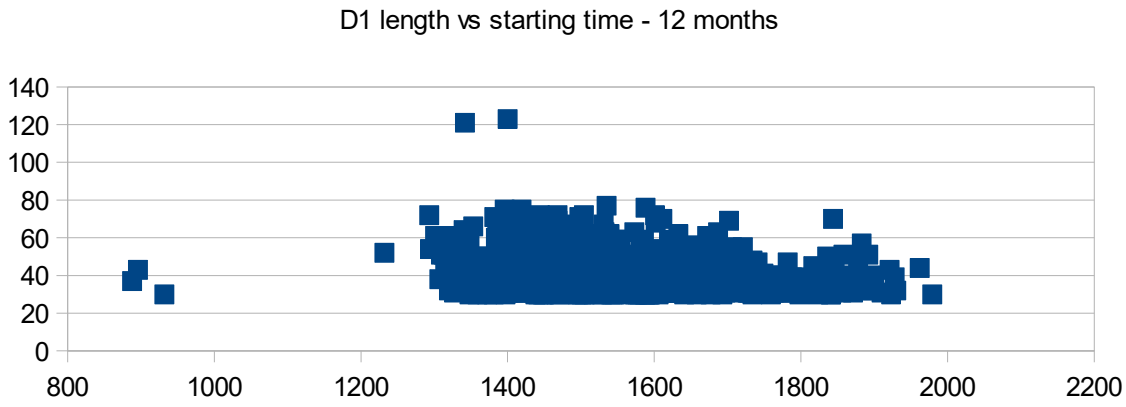


Fig.22 First year - D1 segments length Vs starting time. X axis 1440= Midnight

Note 2) It is quite interesting that the only 3 afternoon values (I'm not used to nap) are inside the same hour, pointing to the sleep "second door".

Note 3) D1 daily total is always fluctuating between 30 and 227 (Fig. 23).

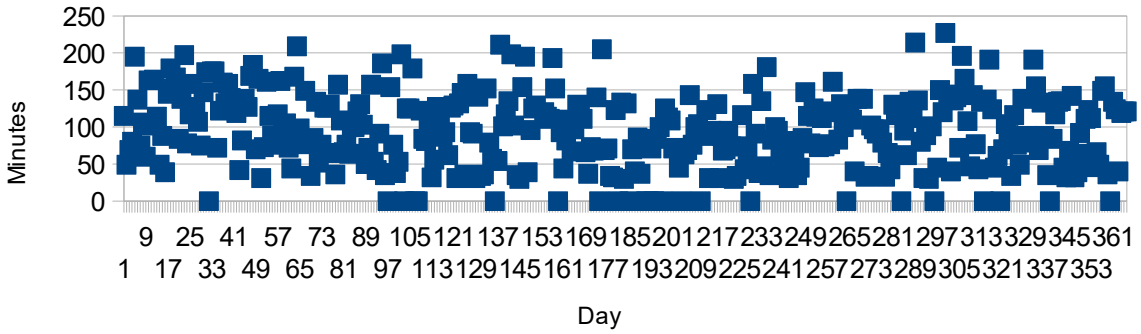


Fig.23 First year - D1 day total

But, if we compute the running mean of 5 days, or search for a fit with a sinusoidal curve, we find some kind of circannual rhythm (Fig.24)

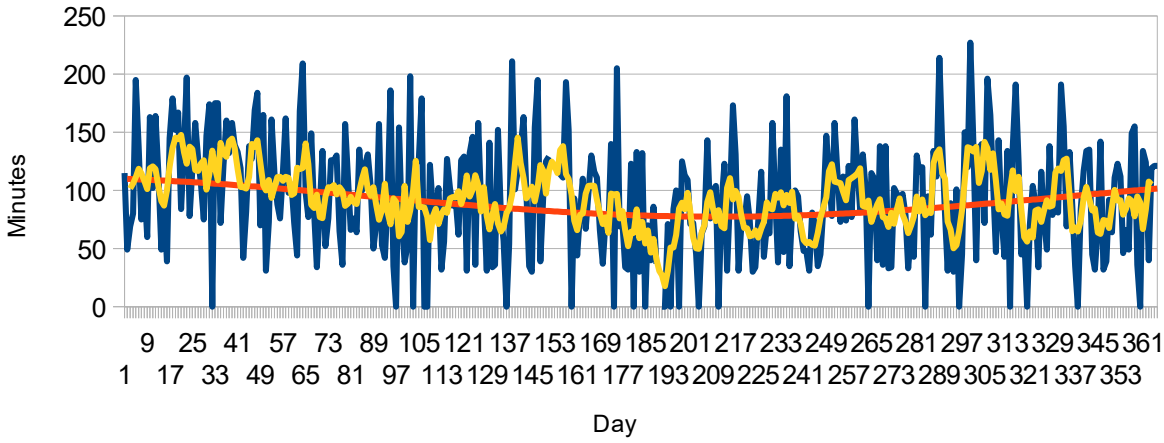


Fig.24 D1 day total data (blue), 5 days total running mean (yellow), annual trend (red).

Since day 1 is December 17, it is suggestive to imagine winter and springtime going down (Winter 21/12 – 20/3, Springtime: 21/3 – 20/6) and summer and fall going up (Summertime: 21/6 – 20/9, Fall: 21/9 - 20/12).
It will be possible to confirm it with the recordings from Phase III (Second year, first semester). If confirmed, it will then be possible to analyse "reactions" as in Fig.12.

Note 4) D2 appears nearly only during bedtime, but it could also be recorded during "wake" period outside bed. Personal experience seems to point to those sequences as drowsiness and the similar.

Note 5) D1 and D2 total daily mean computed over the year is similar, about 100 minutes each. During nights without D1, D2 mean is 50% higher than his annual mean.

Summing up, I think that it is reasonable to make the following research statements as a guidelines for further analysis:

- Some amount of time spent in long zero sequences (with duration more than 30 minutes, D1) during 24 hours is a body requirement. When not reached, it is needed a recovery and that "debt" cumulate over days and must be fulfilled as soon as possible.
- The daily mean total value of long zero sequences has a circannual rhythm, with a peak at the winter beginning and a low at summer beginning, that is possible to forecast.
- Long zero sequences are not possible late morning and late afternoon, defining sleep first and second "doors".
- If the first long zero sequence of the day starts after a time limit, the total of the day will be lower than the required daily mean.
- Alcohol limits the long zero sequences total value (from sleep diary).
- Over that year, I never went to a movie theatre. As far as I remember, it is the strongest

"immobility's inducer" and it would be a nice test for D1 strength against artefacts.

For co-recorded polysomnography

- Long zero sequences of wrist actigraphy are possible only during sleep; start and stop of a long zero sequence relate to some body state.
- Co-recording of actigraphy and polysomnography has been made in the past, mainly for validation of actigraphy parameters. It is a database of data "ready to use" and zero series analysis is available in the standard software of the Actiwatch/MotionWatch recorders. In a polysomnography study, no other methodology allows to correlate two events that are as far as more than two hours, with the precision of one second.

Notes about data recorded during Phase I and II of the project.
During those 12 months, there are 3 nights (152,153,190 – 16/5/16,17/5/16,24/6/16) not recorder due to management mistakes, Those are marked as NA.
3 nights (200,236,245 – 4/7/16, 9/8/16, 18/8/16) are only partially recorded and the partial values are used.

There are 24 nights without a D1 recorded, and 0 has been used for computation (Fig.18). It seems that the 30 minutes criteria used for the first research hypothesis was a lucky case.
Is the distribution of those nights random?

Summer legal time started on Sunday 27/3/2016 until Sunday 30/10/216 (nights 102 to 319). Watch time has been always used for computation.