TDK-dolgozat

Komáromi Krisztina Panna

The effects of processing and analysing methods on describing the scale-free nature of human locomotor activity

Komáromi Krisztina Panna

IV. year molecular bionics engineering BSc.

Supervisors: Dr. Vadai Gergely, Maczák Bálint

University of Szeged, Department of Technical Informatics

Aktigráfiás jelek feldolgozási és elemzési módszereinek hatása az emberi mozgás skálafüggetlenségének jellemzésére

Komáromi Krisztina Panna
IV. évf. molekuláris bionika mérnök BSc.
Témavezetők: Dr. Vadai Gergely, Maczák Bálint
SZTE TTIK Műszaki Informatika Tanszék

Bár az emberi dinamika skálafüggetlen jellege mind a térbeli mozgás, mind a motorikus aktivitás esetében tetten érhető, a különböző adathalmazok statisztikai eloszlásait vizsgálva a hatványtörvényen túl azonban számos egyéb egymásnak ellentmondó eredmény is született. Az emberi motorikus aktivitás mérését gyakran alkalmazzák orvosi célú (pl. alváskutatás, pszichiátria) kutatásokban is, amit egy ún. aktigráffal végeznek, mely a csukló gyorsulása alapján számszerűsíti az alany aktivitását. Ennek megvalósítása azonban gyártónkként eltérő és többnyire nem publikus, ami miatt sokszor eltérő módokon előállított aktivitásjeleket vizsgálnak a mögöttes módszertan kellő jellemzése nélkül. Továbbá, az aktivitásadatok statisztikai elemzési módszerei sem meghatározottak a szakirodalomban. Például eltérő lehet az, hogy az aktivitásjel aktív és passzív szakaszait milyen küszöbszinttel különítik el, ezek hosszait milyen, az eloszlást jellemző függvényen (valószínűségi sűrűségfüggvény, túlélési függvény) keresztül vizsgálják, arra hogyan (*least squares, maximum likelihood*), mekkora tartományon, milyen eloszlásokat illesztenek és miképp választják ki a legmegfelelőbbet. Ezzel szemben a kutatócsoport korábban megmutatta, hogy a csuklón mért gyorsulásjelek, és az azokból számolt aktivitásjelek spektruma is egy univerzális, spektrális skálafüggetlenséget jelentő, $1/f$ zaj alapú karakterisztikát követ az aktigráfiás módszertantól függetlenül.

Ezek alapján célom a kutatócsoport korábban is vizsgált gyorsulásjeleiből az eltérő módszerekkel meghatározott aktivitásjelek statisztikai jellemzőinek összevetése és ezek alapján az okozott eltérések vizsgálata volt. Ezen felül pedig a legfontosabb kapcsolódó publikációk eltérő modellillesztési megközelítéseit összevetve is megmutatni, hogy azok milyen különbségeket okoznak. Eredményeim azt mutatják, hogy az azonos mérésekből eltérő, a szakirodalomban fellelhető feldolgozási módszerekkel meghatározott aktív és passzív szakaszok eloszlása nagymértékben különbözik. Továbbá, a különböző kutatások eltérő modellillesztési módszereit a saját adatsorainkon alkalmazva, a választott megközelítés és paraméterek függvényében is más-más eredményeket kaptunk, ami megerősíti azt, hogy az emberi aktivitás skálafüggetlensége robusztusabban vizsgálható frekvenciatartományban.

The effects of processing and analysing methods on describing the scale-free nature of human locomotor activity

Komáromi Krisztina Panna
IV. year molecular bionics engineering BSc.
Supervisors: Dr. Vadai Gergely, Maczák Bálint
University of Szeged, Department of Technical Informatics

Although the scale-free nature of human dynamics is apparent in both spatial movement and locomotor activity, several contradictory models have emerged in addition to the power law from analysing the statistical distributions of different datasets The measurement of locomotor activity is often used in medical research (e. g. somnology, psychiatry), carried out with a so-called actigraph, which quantifies the subject's activity based on the acceleration of the wrist. However, its execution varies widely between manufacturers and is commonly unpublished, which is why differently produced activity signals are often analysed without giving a sufficient description of the underlying methodology. Furthermore, the statistical analytic methods of the activity data are also unstandardized in the literature. For example, differences may occur in the threshold, that is used for separating the active and passive periods, the distribution function (probability density function, complementary cumulative distribution function) used for examining their lengths, the fitting method (least squares, maximum likelihood) and its range, the kind of distributions fitted, and the method used for determining the most suitable one of them. In contrast, the research group showed previously, that the spectra of both acceleration signals, measured on the wrist, and activity signals derived from them follow universal $1/f$ noise-based characteristics, which represents the spectral scale-free property, regardless of the actigraphic methodology.

On this basis, my goal was to compare the statistical characteristics of the activity signals calculated with different activity determination methods from the acceleration dataset, previously analysed by the research group to investigate differences they may cause. In addition, by comparing the different model-fitting approaches of the most important related works, I aimed to show their influence. My results show, that the distributions of the active and passive periods calculated with the processing methods prevalent in literature from the same measurements differ greatly. Furthermore, using the different model-fitting techniques of the related works on our datasets produced varying results, depending on the chosen approach and parameters. This reinforces us that the scale-free nature of human activity can be examined more robustly in the frequency domain.

Table of contents

1. Introduction

Human dynamics exhibit scale-free characteristics in both spatial movement data and locomotor activity measurements. However, conflicting models have emerged from the statistical analyses of different datasets beyond the power law distribution. Locomotor activity is often measured with actigraphs in medical research, but the differences in their methods for quantifying activity from 3-axis acceleration is very inconsistent and often unpublished, making it hard to reproduce. This difference in method also seems to affect their statistical analyses. In addition, the other aspects of the examinations, done with statistical methods are also not standardized and this may be related to the varying results in the field. The threshold, which is used to distinguish between the active and passive periods, seems to be unstandardized, along with the type of distribution to be calculated from these periods, which is fitted with the heavy-tailed models later in the analysis. The method and range of the fit also seems inconsistent.

In contrast, the research group has shown that the spectra of both acceleration signals and activity signals derived from them follow universal $1/f$ noise-based characteristics, which represents the spectral scale-free property, regardless of the actigraphic methodology [1]. This raised the issue to find out what exactly caused the difference in the results of the studies, which used the statistical approach. Recently, we demonstrated the differences caused by the different methodologies of processing the data [2]. Since then, we also investigated the problem of commonly used model-fitting methods in the relevant literature. This study aims to show the effects of the different activity determination methods and statistical approaches by comparing the techniques prevalent in literature in a unified framework. The first section of the study contains the theoretical background of the topic, covering the related fields, the mathematical background of the methods and a review of the relevant literature. The next section contains the information about the data acquisition and the activity determination and statistical methods. The last section of the paper aims to demonstrate the acquired results and compare them with the relevant publications, drawing conclusions from them about the effect of the methodological differences. The results show that both the differences in the activity determination methods and the different statistical approaches in the analysis can cause the contradicting results, furthermore, suggest that examining the scale-free nature of human locomotor activity is more robust when analysed in the frequency domain.

2. Theoretical background

2.1. Related fields of human dynamics

2.1.1. Scale invariance

The concept of scale invariance has been a prominent topic of discussion within the scientific community in the last few decades. First, the scale invariance of networks has been investigated. A scale-free networks degree distribution follows a power law, meaning that there are a lot of nodes with few connections, and very few nodes with a lot of connections. Fractals are good examples of scale invariance because they are infinitely complex patterns, that exhibit self-similarity across varying scales. After the examination of networks, the modelling of animal wandering was one of the next steps for the topic of scale invariance. Previously, their movement was described as Brownian motion, which can be imagined as the random movement of particles suspended in a medium, tracing out a trajectory. Regular Brownian motion is when the "jumps" taken by the particle are independent and Fractional Brownian motion is when the jumps are correlated in time, meaning that the next jump tends to be in the same direction as the previous one was. Brownian motion is a random walk model, as well as the Lévy flights model, where the step size is usually described by power law distributions, or more generally, heavy-tailed distributions. Studies have shown that these were better to model the foraging motion of many living organisms, thus indicating scale-free patterns [3]. This raised the question of whether human motion exhibits the same characteristics, and numerous studies have been conducted based on different approaches to the measured signals and obtained different results. [4] [5].

In case of the frequency domain, the scale independence is addressed as $1/f$ noise, or pink noise, where the power spectral density is inversely proportional to the frequency of the signal. However, the exact mathematical relationship between scale-free characteristics in time and frequency domain is not clearly described yet.

2.1.2. Actigraphy

Human movement can be examined through its mobility and activity. Spatial mobility can be measured using GPS signals, while locomotor activity can be calculated from acceleration signals with the method of actigraphy. It is a non-invasive method often used both in medical research and everyday life.

Actigraphs record raw acceleration data, usually measured on the subject's non-dominant wrist, in up to three axes, then determine the activity from them, the method of which can vary significantly, as investigated by the research group previously [6]. This activity determining method consists of first preprocessing the data, that is essential to remove the Earth's gravity by filtering it out or normalizing, which can be done on all three axes, or the magnitude of all acceleration. Then, an activity metric is used to calculate the activity values from the preprocessed data for every epoch, which are non-overlapping timeslots of equal length [6]. The most commonly used ones in the relevant literature are the Zero Crossing Method (ZCM), which determines the activity values from counting acceleration signal threshold crosses, and the Proportional Integration Method (PIM), which is based on the integration of the acceleration data. The activity values are then separated into active and passive periods using a threshold on the activity signal (which is not the same as the threshold used by the ZCM activity metric, which is used on the acceleration signals) to determine which is which, and then usually the length of these active or passive periods are examined. The process is shown in Figure 1 The other methods will be described in section 3.1.

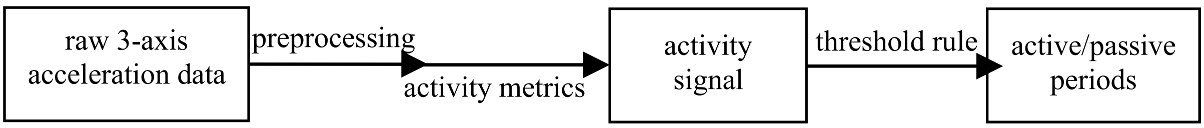

Figure 1. *The process of determining the active and passive period lengths from the raw acceleration data.*

Actigraphic devices are often used in somnology to measure sleep-wake rhythms, because of their many advantages, namely that they provide objective information, document day-to-day variability and the recordings are not altered by laboratory environment compared to other investigation methods [7]. In the psychiatric field, actigraphs can aid in the diagnosis of mental disorders [8]. Both fields utilize statistical analysis in research to draw correlations between healthy and abnormal behaviour from activity recordings. In everyday life, a lot of people have their activity measurements taken by their smartphones or commercially available wearables. However, the exact procedure of quantifying activity based on acceleration varies widely between manufacturers, even in actigraphs used in research and is commonly unpublished, therefore the studies are not reproducible. Still, it is a widely used method where the power law scaling or other heavy-tailed distributions and the associated scale-free nature of human movement is apparent in the distribution of active and passive period lengths, as various studies have investigated it using statistical approaches, with varying results [9 -17].

2.2. Mathematical background

2.2.1. Distribution functions

For the model-fitting, the datasets' Probability Density Function (PDF) and Complementary Cumulative Distribution Function (CCDF) has been calculated, since both are common in the relevant literature. The PDF is a function of a random variable and the relative likelihood of measuring the particular values. We can obtain the PDF if we derivate the Cumulative Distribution Function (CDF), as it can be seen in Eq. 1 and Eq. 2.

$$F_X(x) = P\,(X < x) \tag{1}$$

$$f(x) = F'(x) \tag{2}$$

The properties of the PDF are:

$$f(x) \geq 0 \tag{3}$$

$$P(a < X < b) = \int_a^b f(x)dx \tag{4}$$

Therefore:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{5}$$

To determine the PDF in practice, first the histogram of the data is calculated, then each frequency is divided by the total number of data points, and the associated bin width. In the case of our data instead of linear binning, where each bin is the same width, we use logarithmic binning, which means that the bins will have the same width logarithmically. This is important, because the range of the values and their probabilities spread over a wide range, and the tail of the distribution would be noisy otherwise, which affects the results, because the distribution of passive periods is mainly fitted with heavy-tail distributions. The CCDF is determined by subtracting the CDF from 1 (Eq. 6).

$$\bar{F}_X = 1 - F_X \tag{6}$$

The CDF defines the probability that a random value will be less than x, while the CCDF describes the opposite, how often the variable is above or equal to a particular value. Both the PDF and the CCDF are used in analysing the scale-free nature of human motion. In our case, the human locomotor activity is scale-free, if the PDF or the CCDF of it follow a power law distribution, with the exponent of α in the case of the CCDF, and exponent of α+1 for the PDF.

2.2.2. Model-fitting

The model-fitting methods mentioned in this study were the Maximum Likelihood Estimation (MLE) and the Least Squares method (LS). Maximum Likelihood Estimation is a statistical method used for estimating the parameters of a model by maximizing the likelihood that the assumed model results in the observed data. The Least Squares method is another statistical approach used for estimating the parameters of a model, but it operates by minimizing the total sum of squared error (SSE) between the observed and the fitted values. The squared difference is used since the deviation could be both negative and positive. Our data has a logarithmic scale, therefore when fitting with this method, the tail of the distribution of the data will not be taken into account with the same weight as the initial part of the range, because the values are much smaller there. To overcome this problem in Least Square fitting either different weights can be applied, or the log-transformed forms of the models can be fitted to the data (which is what we chose to follow in this study).

When determining the best fitting models the following goodness of fit metrics found in the relevant literature can be used: Likelihood Ratio Test (LRT), Sum of Squared Error (SSE), Reduced $\chi 2$, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). It is important to choose a goodness of fit method that accounts for the logarithmic scale of the data, otherwise the determined best-fitting model may not fit the tail of the distribution that well. The LRT compares the goodness of fit of two models based on their likelihood ratio. The SSE is calculated by subtracting the fitted y values from the original y values, squaring the differences and adding them together. This is the value that the LS fitting method minimizes.

$$SSE = \sum_{i=1}^{N} (y_{data}(i) - y_{fit}(i))^2 \tag{7}$$

The reduced chi square is calculated by dividing each squared errors with the sum of the fitted y and the original y values, then adding them together and divide the sum by the number of degrees of freedom. [16]

$$Reduced\ \chi^2 = \frac{1}{degrees\ of\ freedom} \sum_{i=1}^{N} (y_{data}(i) - y_{fit}(i))^2 / (y_{data}(i) + y_{fit}(i)) \tag{8}$$

The AIC is calculated by multiplying k (the number of parameters in the model) with 2 and subtracting L (the loglikelihood function) multiplied by 2 from it. The BIC is calculated by multiplying k with the logarithm of N (the number of data points) and subtracting 2 times L from it. [16]

$$AIC = 2k - 2L \tag{9}$$

$$BIC = k \ln N - 2L \tag{10}$$

The difference between AIC and BIC is that, while AIC balances between model fit and model complexity, BIC penalizes model complexity and prefers simpler models. The different results in determining the best fit may occur due to these differences.

2.3. Literature review

We collected 10 important publications in the field of statistical analyses of human actigraphy signals from the past 20 years and analysed their differences focusing on their activity determination methods and statistical approaches. This review will concentrate only on the model they fit to their data in the end, not all the models they tried to fit.

From the five studies which examined activity determined with the previously mentioned ZCM activity metric, all but one fitted power law to the distribution of passive periods, the exception being the exponential function [13]. For the active periods three of them fitted stretched exponential [16] [14] [9] and the other two fitted a power law distribution [13] [17] (with one of them being the one that was also distinct from the rest when examining the passive periods). Four of the studies used the PIM activity metric, and to the passive periods they fit power law, power law for only the initial range and truncated power law while to the active periods, power law for only the initial range, truncated power law and lognormal, which suggests that it is not only the commonly used activity metrics that have a significant impact on the results of the fit. However, the different preprocessing pairings may also have an influence, but specific conclusions could not be drawn, because it was not possible to determine the preprocessing used in several of the studies, so only a general examination has been conducted.

The activity values are separated with a threshold to determine the active and passive periods, as stated previously. The value of this threshold was also not consistent across the articles. Various threshold rules occurred, using the 1, 10, 70 or 100 percent of the mean activity value or the mean of the non-zero activity values. The three studies which used thresholds greater than the mean of the overall activity [9] [15] [14] fitted power law or truncated power law to the passive periods and stretched exponential or truncated power law to the active ones. The seven articles using thresholds smaller than or equal to the mean [13] [11] [12] [16] [10] [18] [17] fitted power law, power law for only the initial range, exponential or lognormal to the passive periods, which seems very diverse, and fitted power law, power law for only the initial range, truncated power law, lognormal or stretched exponential to the active periods, which seems even more contrasting. It can be seen that from the studies who fit distributions other than power law or truncated power law to passive periods all used thresholds less than or equal to the mean of the period lengths. In section 4.4 the change in the shape of the curve as a result of modification of the threshold will be examined further. In the studies where the active periods were fitted with some kind of power law, the threshold also seems to be below the mean of the period lengths.

The fitting to distributions of active and passive periods was determined individually in six articles, and on pooled data (which is concatenating every person's period lengths together) in four of them. A conclusion could not be drawn about whether this affects the result of the fitting due to the conflicting results. The majority of the articles calculated the CCDF of the active and passive period lengths, but two of them seemed to only [12], or in addition [18] calculate the PDF too.

For the model-fitting some articles used Maximum Likelihood Estimation (MLE), and some of them used the Least Square (LS) method. Setting *xmin* and in the case of LS, *xmax* values were also common, six of the articles used some kind of boundary for the range of the fit, but these were not uniform. There were also a lot of different approaches to determining the best of the fitted models, the most common were the $\chi 2$, AIC, BIC and LRT. However, some of the studies only fit one model, in their case the goodness of fit metric is sometimes used to determine the best threshold rule [16] or to determine the dependency of the distributions on the threshold or the data resolution (epoch length) [14].

It seems that while some conclusions can be drawn regarding the effect of the different methods, the problem seems to be too complex and there is not sufficient information given about the specifications of each method used. An attempt was made to reproduce the studies to further understand this problem. In the cases where the attempt was successful, my results coincided with the ones in the articles. A primitive reproduction of the articles can be found in section 4.7, where only the activity metric, the preprocessing method and the threshold rule is taken into consideration because a lot of the crucial steps were missing from the articles.

2.4. Objective of research

Our objective was to find the underlying rules behind the contradicting results and show that the results of the statistical approach to the analysis of human locomotor activity is very dependent on the methods, which are not standardized.

The research group recently showed that the Power Spectral Density (PSD) of the activity signals, and even the raw activity signals follow universal $1/f$ noise-based characteristic. They demonstrated that this spectral scale-free nature is apparent, regardless of the activity determination method used. This suggests that the spectral analysis seems to be more robust in examining scale independency than the statistical analysis mentioned above, because of its inconsistencies.

Our objective was to show the differences the various activity determination methods and the different threshold rules made during the different statistical analyses.

In addition, I intended to examine the effect of the different model-fitting methods, used by the mentioned articles, to get closer to how these methodological differences contribute to the aforementioned contradictions.

3. Materials and methods

3.1. Data acquisiton, activity calculation

For the analysis the research group's data has been used, which consists of 10-day-long raw acceleration recordings of 42 healthy, free-living individuals, which they also used in their previous works [6] [1] [2]. The signals were measured at the sampling rate of 10 Hz in the ±8 g range with an actigraphic device on their non-dominant wrist, that was specially developed to store raw triaxial acceleration data.

The research group's objective was to compare the different activity determination methods, which is why they measured raw acceleration data. Then they determined the different activity values from the acceleration data using different preprocessing methods and activity metrics.

For determining the activity values, first the acceleration data was preprocessed to remove the gravity of Earth. This can be done in various ways, for example, by calculating the unfiltered magnitude of acceleration (UFM, i.e., the length of the acceleration vectors) from the unfiltered acceleration values of the x, y, and z axes (UFX, UFY, UFZ, respectively). Then, the effect of Earth's gravity (g) can be eliminated by calculating the absolute distance of the magnitude values from 1 g, which results in the unfiltered normalised magnitude of acceleration (UFNM). The elimination of g can be achieved with a bandpass filter, too, which filters out low frequency components, and high-frequency noise, such as components related to involuntary movements. The filtering can be executed in two ways. Firstly, the filter can be applied on the magnitude of acceleration data (UFM), resulting in the postfiltered magnitude of acceleration (FMpost). Secondly, the filter can be applied to the per-axis acceleration (UFX, UFY, UFZ) prior to magnitude calculation, resulting in the filtered acceleration values along the x, y, and z axes (FX, FY, FZ), from which the prefiltered magnitude data can be calculated (FMpre) After preprocessing, the data was split into consecutive timeslots of equal length (epochs),which was 60 seconds in our case (while the 60 second epochs are the most common in the literature, other values occur, which is also a difference in activity determination) and an activity value was determined for each epoch using an activity metric. The research group previously collected 7 typical metrics that operate on significantly different principles. One of the simplest metrics, that I already mentioned in sections 2.1.2 and 2.3, (PIM – Proportional Integration Method) is based on the integration of the acceleration data. Other classical metrics rely on the threshold intersections of the acceleration data: whereas Zero Crossing Mode (ZCM), also mentioned previously, counts the threshold crosses, Time Above Threshold (TAT) measures the time when the data exceeds the threshold. In addition, there are metrics that averages the acceleration data: while the Euclidean Norm Minus One (ENMO) utilizes a special averaging rule, the High-pass Filtered Euclidean Norm (HFEN) requires specially preprocessed acceleration data. Beyond these, metrics relying on the standard deviation (MAD – Mean Amplitude Deviation) of the magnitude of acceleration or variance (AI – Activity Index) of the per-axis data also exist. Thus, there are certain metrics that require acceleration data preprocessed in a strict way, while others can also be applied to differently preprocessed acceleration signals. The details about the limitations, compatibility, and proper combinations of preprocessing techniques and activity metrics can be found in the research group's previous work [6]. To identify activity signals determined in a given way, I will use the following notation in line with the previous works of the research group: I denote the activity metric as an operation, and the preprocessing method as its argument (e.g., ZCM(FMpost).

After the activity values were calculated, a threshold, which is usually defined as some percentage of the mean of the overall activity, as shown before in section 2.1.2, was used to define the active and passive period lengths. The activity values, calculated for every one-minute epoch were considered as ones when above the threshold and as zeros when below. Then the length of the consecutive ones or zeros were considered the active and passive period lengths, respectively (in our case, the number of ones/zeros equal the length in minutes because of the 60 second epochs). A different theoretical approach is to consider consecutive non-zero activity values as events of a given duration and the time elapsed between them as waiting times; however, similar models could be fitted to the distributions in this sense, too [13]. Therefore, a technical analogy can be instituted between this and the former approach if the threshold level is chosen to be negligibly small (i.e., 1% of the mean(a)).

3.2. Statistical methods

Once the active and passive period lengths have been determined, the two numerical approaches for examining them that are covered in this study, because of how common they are in the relevant literature, are the estimation of their Probability Density Function (PDF) or their Complementary Cumulative Distribution Function (CCDF). It is not obvious which one is better to use, because both have various advantages and disadvantages. To determine the PDF, the data is split into bins, the number of which may be problematic to choose. If it is too high, there will be a lot of bins with a negligible number of samples, and the noise in the tail gets more defined, and if it is too low, the PDF will have low resolution. The bin widths can also be linear or logarithmic, of which the latter can reduce noise in the tail, because there will be fewer bins on the end of the axis range. The problem of binning can be avoided if the CCDF is used, however, the power law scaling is more prominent in the PDF. The CCDF deviates from the ideal line of the power law scaling at longer periods. This truncation may be the result of the finite size of the sample, which is a known phenomenon [19] [20]. Both the PDF and the CCDF can be estimated individually or pooling the individuals' data together, which is concatenating the active/passive period lengths of all individuals and calculating the distribution function from them. The mean activity value, used for the threshold can be calculated one by one for each measurement, or altogether. In the upcoming figures and tables, the approach of pooling the individual measurements together is used if not stated otherwise.

Another considerable change is the rescaling of the distribution function, which means dividing the values with their mean, or some percent of the mean. It can come up when comparing the CCDFs of data covering different ranges, e.g., when comparing human and animal activity [14].

For the model-fitting two methods were used, the Maximum Likelihood Estimation and the Least Square Estimation, both were mentioned previously in sections 2.2 and 2.3. For MLE, a Python package was used [21], which is able to display the PDF and CCDF of the data, and fit power law, exponential, lognormal, truncated power law and stretched exponential models. It can also estimate the best fitting model by the Likelihood Ratio Test. The LS method is carried out with the SciPy curvefit function in Python and could fit power law, exponential, lognormal, truncated power law and Weibull (stretched exponential) models to the data with some restrictions applied to the parameters. The range of the fit in MLE could be modified with the *xmin*, and in case of the LS method both the *xmin* and *xmax* of the range could be set. This marks a time interval excluding the active/passive periods shorter than *xmin* or longer than *xmax* from the fit. The best fitted model can be calculated with various goodness of fit measures. The ones used in the study are sum of squared error, reduced chi square, Akaike Information Criterion, Bayesian Information Criterion and the Python package's Likelihood Ratio Test.

4. Results

The results of the analyses of the methodological differences in the activity determining methods and in the statistical analysis are shown in the next section. Since the problem is very complex, due to the many parameters that can vary, I will introduce a standard setting, based on the most common methods and only examine the effect of one variable at a time. This will be:

preprocessing:	*FMpost*,
activity metric:	*ZCM*,
threshold:	*100% of mean*,
PDF/CCDF	*CCDF*,
bin number (in case of the PDF):	*100 bins*,
individual/pooled:	*pooled*,
rescaled/non-rescaled:	*non-rescaled*,
fitting range:	*xmin = 1, xmax = maximum of dataset*,
method of fit:	*MLE*,
goodness of fit metric:	*LRT*.

When using the MLE method, the fitted models are power law, truncated power law, exponential, stretched exponential and lognormal. When using the LS method, the fitted models in the case of the CCDF are power law, truncated power law, exponential, stretched exponential and lognormal, and in the case of the PDF, power law, truncated power law, exponential and lognormal. The effects are examined in the order in which the results rely on each other.

4.1. Effect of activity determination

The first step of analysing the scale-free nature of human locomotor activity is determining the activity from the raw acceleration data. There are numerous methods found in the literature, as shown in section 3.1. and their differences have been investigated by the research group in a previous study [6]. My results indicated that the preprocessing method and the activity metric affected the shape of the CCDF and PDF of the pooled passive periods, as shown on Figure 2a and 2b. I found that in the case of ZCM activity metric the preprocessing did not cause a major difference in the shape of the distribution, so only one preprocessing method can be seen on the figure. In the case of the different preprocessing methods used with TAT, there were only minor differences, the general shape was the same, so also only one of them is displayed on the figure. In the case of PIM and MAD activity metrics, only the preprocessing pairings with the most extreme shapes are displayed on Figure 2a and 2b. It can be seen on Figure 2c and 2d that in the case of the PIM activity metric, those were the UFNM (but could have been the FMpost too, they seem very similar) and the FX. The most commonly used activity metrics in the literature were ZCM and PIM, in which case only the preprocessing of PIM could affects the shape minimally, but there is not as much of a difference as, for example with AI (which can be due to the difference in the purpose of the indicator [22]), or between the differently preprocessed MAD metrics. However, in the case of the active periods, the shape of the curves is roughly the same across all different activity determination methods, as shown on Figure 2e and 2d. This suggest that the activity determination method affects the distribution of the passive periods more, and in that case, the differences are significant. In section 4.5 in Figure 9, the effect of the different preproceccings can be seen, I discuss them there, because their differences come out more clearly when seeing the different models the MLE estimates as the best on them.

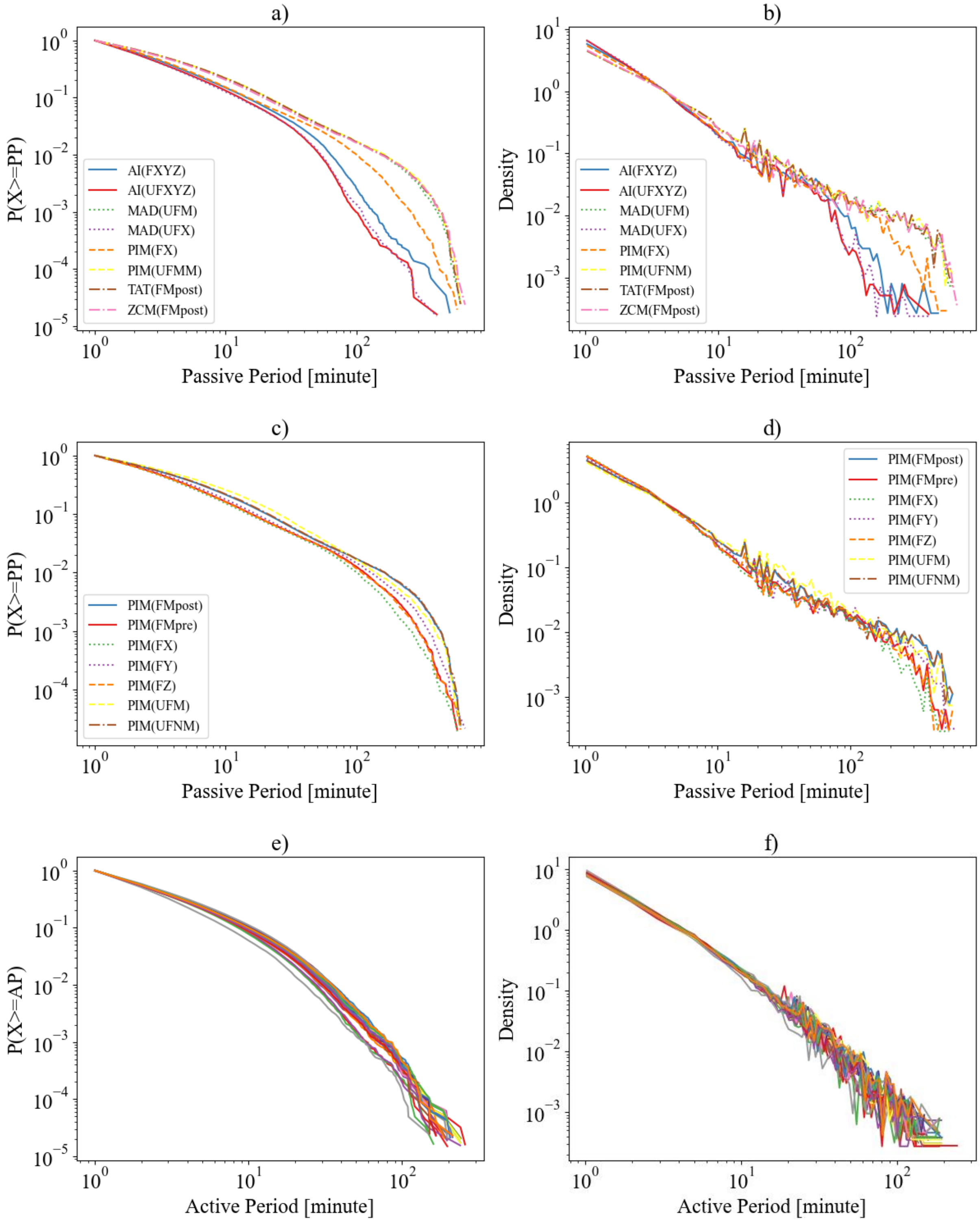

***Figure 2. a**, The pooled, non-rescaled CCDF of passive periods for different types of activity values, that represent the found differences, using a threshold value of average of length. **b**, the same for the PDF using 100 bins. **c**, The pooled, non-rescaled CCDF of passive periods for activity values determined with every preprocessing paired with the PIM metric, using a threshold value of average of length. **d**, the same for the PDF using 100 bins. **e**, The pooled, non-rescaled CCDF of active periods for every type of activity value, using a threshold value of average of length. f, the same for the PDF using 100 bins.*

4.2. Effect of pooling

The best fitting model's parameters can be determined with calculating the average from the parameter fits based on the distribution function of data measured for each individual. It can also be obtained with fitting the distribution of the pooled active or passive periods. These two methods could affect the results, because as seen on Figure 3, the curves are different for the individual and the pooled data. The maximum x value of the pooled curve is the same as the longest individual curve's maximum x value, while the minimum y value is smaller, because the pooled has various different period lengths. This effect can be avoided if the $xmax$ of the fitting is smaller than or equal to the shortest individual curve's maximal x value, because up to that value the pooled curve follows the individual ones, and only the tail is bending down, which can affect the fitting

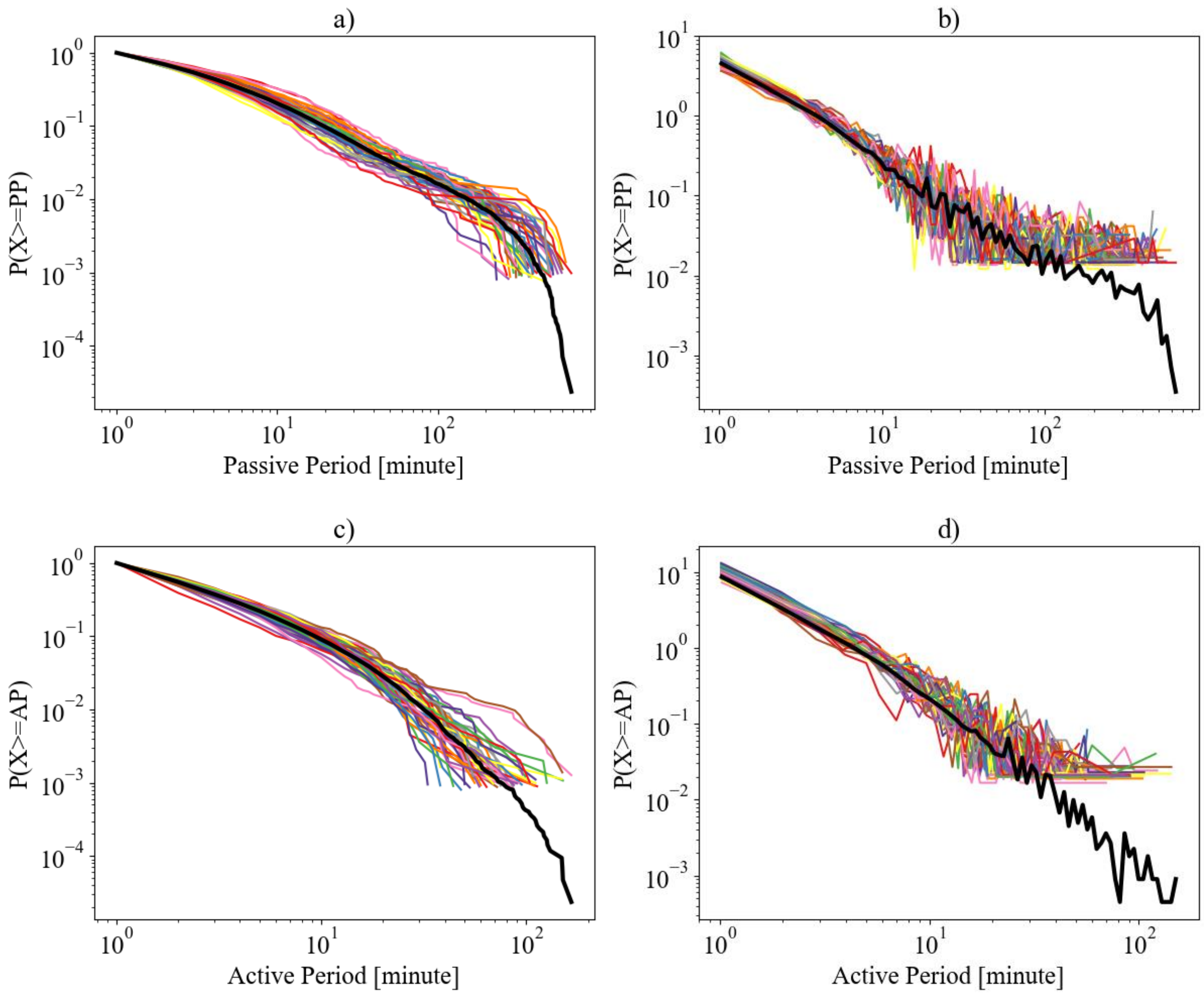

Figure 3. ***a***, *The non-rescaled CCDF of the pooled and individual passive periods determined by the ZCM(FMpost) activity metric, using a threshold value of average of length.* ***b***, *the same for the PDF using 100 bins.* ***c***, *The non-rescaled CCDF of the pooled and individual active periods determined by the ZCM(FMpost) activity metric, using a threshold value of average of length.* ***d***, *the same for the PDF using 100 bins*

4.3. Effect of the number of bins

When determining the PDF, the number of bins used when calculating the histogram can also affect the shape of the curve. As seen on Figure 4., where the more number of bins used, the bluer the curve is, the differences can be in the noisiness and the slope of the curve. The slope can get more steep at the smaller values as the number of bins increases, because the first few bins always have the same samples in them (the smallest periods are always 1, 2, 3, etc. minutes, while the longest ones are always different because of the logarithmic scale of the data), while the bin width is decreasing (by which the number of samples are getting divided), so their probability is increasing. The noisiness that comes with using more bins at the larger values may be caused by the tail being covered with more bins containing less and less samples. When fitting a power law model to the PDF, the α (parameter of the power law model, the slope of the distribution on log-transformed axes) may be affected by the number of bins used.

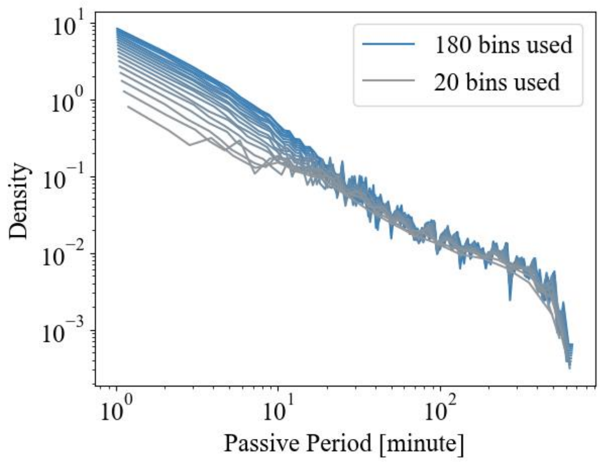

Figure 4. *The pooled PDF of passive periods for different numbers of bins, using a threshold value of average of length and determined with the ZCM(FMpost) activity metric*

4.4. Effect of the threshold

The threshold used for determining the active and passive period lengths can also affect the shape of the PDF and CCDF.

On Figure 5 the activity signals (ZCM(FMpost)) of an individual's first 12 hours of the day can be seen. In their case, when the threshold is determined as the mean of their whole (10-day-long) sample, short active periods can be seen from midnight to 6:30 AM, possibly when they were still asleep. However, with the threshold set to 150% of the overall mean, those short activity spikes while sleeping do not count as active. Therefore, when examining overall human activity, this threshold rule could be sensitive and needs to be adjusted according to what we want to examine. However, in the model-fitting phase, the points belonging to short period lengths could be left out of the range of the fit. An other solution for this problem could be to analyse the sleeping and awake phases separately, determining different thresholds for each, as some studies have done [15] [17]. Figure 6 shows the effect of modifying the threshold on the CCDF and PDF of active and passive periods, the bigger it is, the more blue the curve becomes. The smaller it is, the more the longer passive periods (e.g., sleeping) will be divided by the previously mentioned short activity periods. This results in the passive curve getting shorter and the shape of it changing, while the active curve is getting a little longer. With the threshold set to 150% of the mean, which was approximately the mean of non-zero values regarding our datasets, the maximal length of the passive periods tops out at ~700 minutes, or 12 hours.

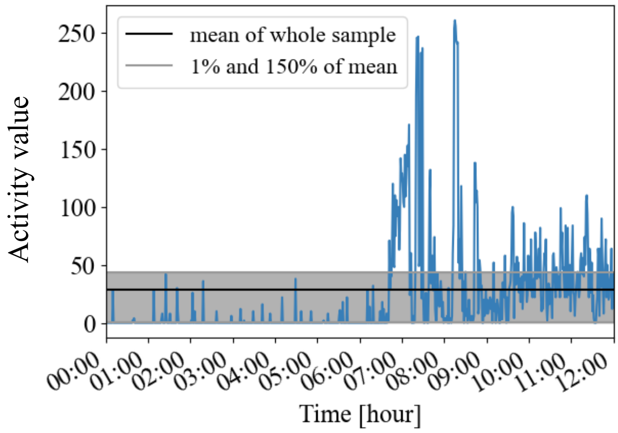

Figure 5. *The first 12 hours of an individual's activity values, determined with the ZCM(FMpost) activity metric. The black horizontal line represents the 100% of the mean of the individual's entire dataset, the grey horizontal lines represent the 1% and 150% of it.*

Figure 7 shows the effect the different threshold rules have paired with specific activity determination methods and *xmins*. As the threshold increases from relatively small to approximately the mean of non-zero passive periods, the best fit determined with MLE becomes the lognormal with almost every one of the various methods in the case of *xmin* = 1 minute. When *xmin* = 10 minutes, at the 10% of the mean value of the threshold, in the case of half of the different activity metrics the truncated power law wins, while in the other half, the stretched exponential seems the best. If the threshold is increased, the truncated power law wins in the majority of the cases. However, if the threshold is increased to the 150% of the mean, more of the best fits become lognormal, making the ratio of the three fits approximately 70-30-0 (truncated power law – lognormal – stretched exponential). There is a significant difference in the composition of the table in the case of the different *xmins*, this means that the *xmin* also plays a role in the determination of the best fit.

Altogether, the threshold rule seems to affect the shape of the distribution, as well as the length of it. This result is consistent with what we have seen earlier, that the studies who determined the best fit as, or only fit a model other than power law or truncated power law used a threshold smaller than or equal to the mean of their dataset, because when a higher threshold is used in our case, the distribution seems more like the truncated power law.

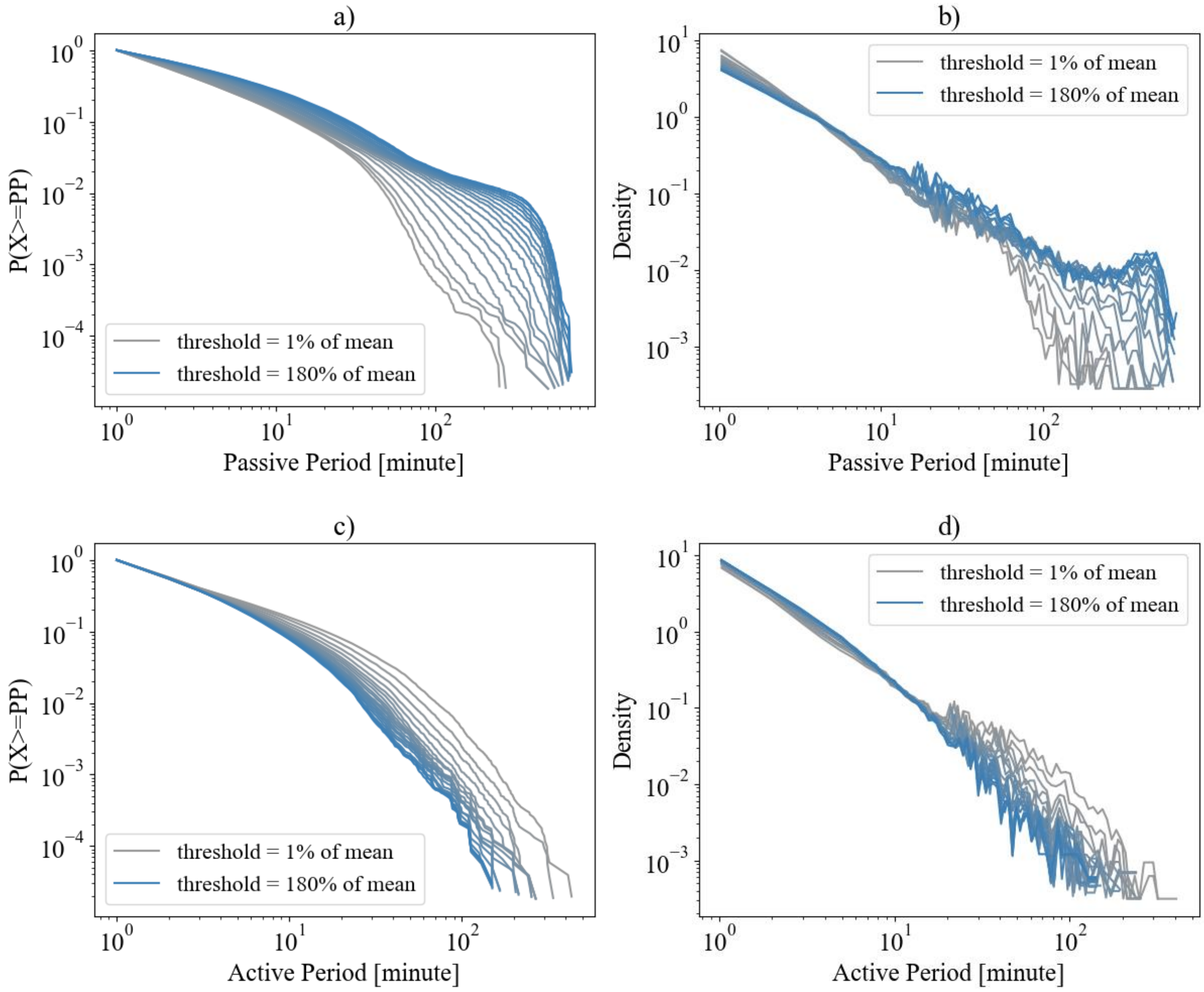

***Figure 6. a**, The pooled, non-rescaled CCDF of passive periods for different types of threshold rules, determined with the ZCM(FMpost) activity metric. **b**, the same for the PDF using 100 bins. **c**, The pooled, non-rescaled CCDF of active periods for different types of threshold rules, determined with the ZCM(FMpost) activity metric. **d**, the same for the PDF using 100 bins*

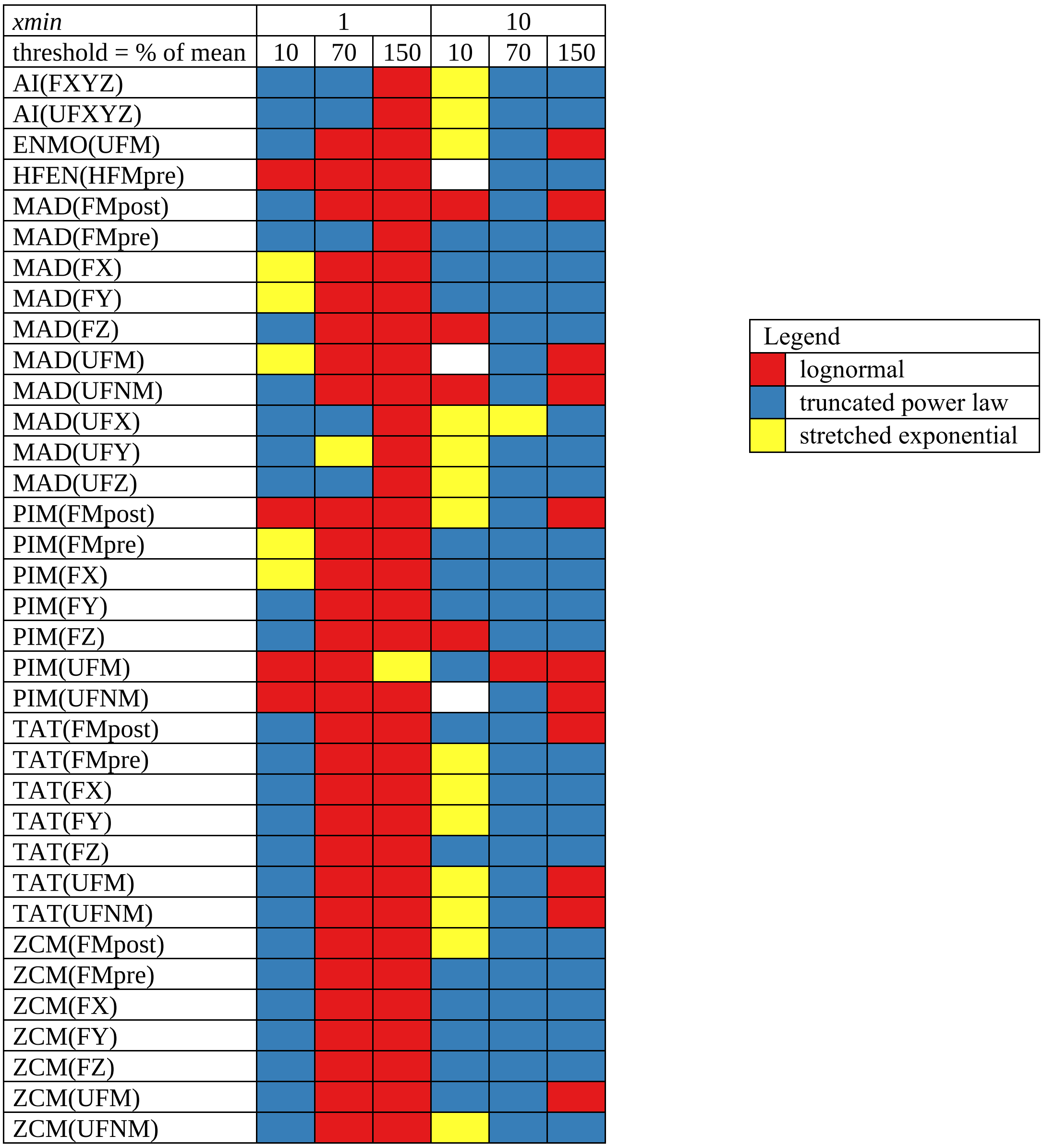

Figure 7. *The best fit determined by the MLE method and the LRT goodness of fit metric on the pooled, non-rescaled CCDF of passive periods, in the case of the different activity determination methods, specific xmins (1 and 10), and specific threshold rules (10%, 70%, 150% of mean). The red cells represent lognormal, the blue cells truncated power law, and the yellow cells stretched exponential. In the case of white cells, a fit could not be determined because the distribution curve was too short and there was not enough data to determine it.*

4.5. Effect of the *xmin* of the fit

Once the PDF or the CCDF of the active/passive period lengths have been estimated, the results can still vary based on the way the fit is performed. Figure 9 contains the best fit according to the Likelihood Ratio Test calculated by the `powerlaw` package [21], determined for every activity determination method and specific *xmins*. It can be seen that when fitting to the whole range of the data the best model seems to be the lognormal, regardless of the activity determination method. However, as the ***xmin*** is increased, the prevalent model starts to become the truncated power law, as depicted on Figure 8, with the method seemingly affecting when it switches from one to the other. When only fitting the tail of the distribution, approximately half of the results are stretched exponential, the other half being truncated power law. The thickly framed parts of Figure 9 mark the groups of the activity determination methods where the preprocessing was the same. The table shows that the ones with the same preprocessing behave in roughly the same way when changing the *xmin* of the fit. This means that the preprocessing also has an effect on the results. The research group previously investigated the problem of the difference in the activity determination methods and found that the preprocessing has the more significant effect in it [6]. In conclusion, the *xmin* of the fit greatly affects the results, because when the fitting on the whole range, the predominantly obtained fit is the lognormal (with the standard settings that I introduced at the beginning of section 4.), while with increasing the *xmin*, thus fitting the tail of the curve more and more, the truncated power law takes its place. When fitting only the tail, half the results seem to be truncated power law still, however the other half seem to fit stretched exponential.

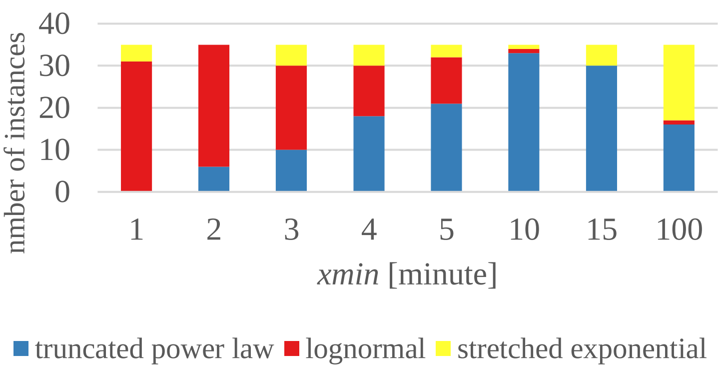

Figure 8. *The distribution of the estimated best fitting model in the case of varying xmins. The red colour represents the lognormal distribution, the blue the truncated power law distribution and the yellow the stretched exponential distribution.*

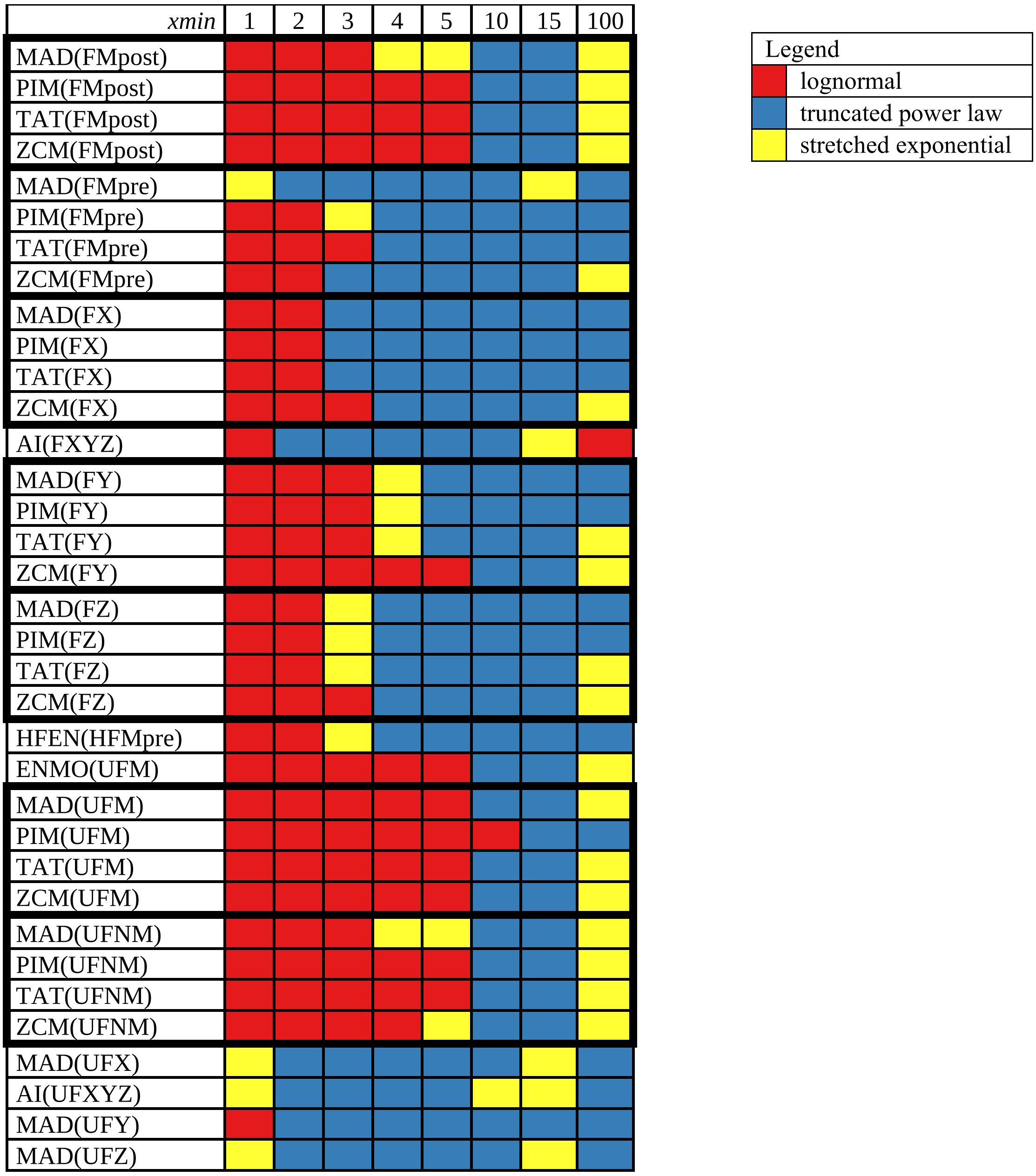

Figure 9 *The best fit determined by the MLE method and the LRT goodness of fit metric on the pooled, non-rescaled CCDF of passive periods, in the case of specific activity determination methods and specific xmins. The activity determination methods are grouped together, indicated with thick black frames, based on their preprocessing methods. The red cells represent lognormal, the blue cells truncated power law, and the yellow cells stretched exponential.*

4.6. Effect of the method of fit

For exploring the differences made by the different fitting methods, the MLE and LS method have been utilized, since both are prevalent in the literature. With the LS method it was easier to account for the logarithmic range of the data, fitting the log transformation of the different models, thus obtaining better fitting on the tail of the curve. This difference can especially be seen on the power law and exponential models on Figure 10a and 10b. All of the fitting was done with *xmin* = 8 minutes, in the case of the MLE (Figure 10a) the fit was estimated on the data, while in the case of the LS, the fitting was done on the CCDF (Figure 10b) or the PDF (Figure 10c) of the data. The MLE estimated the lognormal distribution as the best fit, while in the case of the LS method the truncated power law was the best, both when fitting to the PDF and the CCDF. However, when the *xmin* is modified the LS does not give the same result for the PDF and the CCDF, and the standard setting can also be modified, so that all three of them give different results. However, even if all three of them gave exactly the same fit to the results, when choosing the best of them we can apply different goodness of fit metrics, as seen on Figure 12.

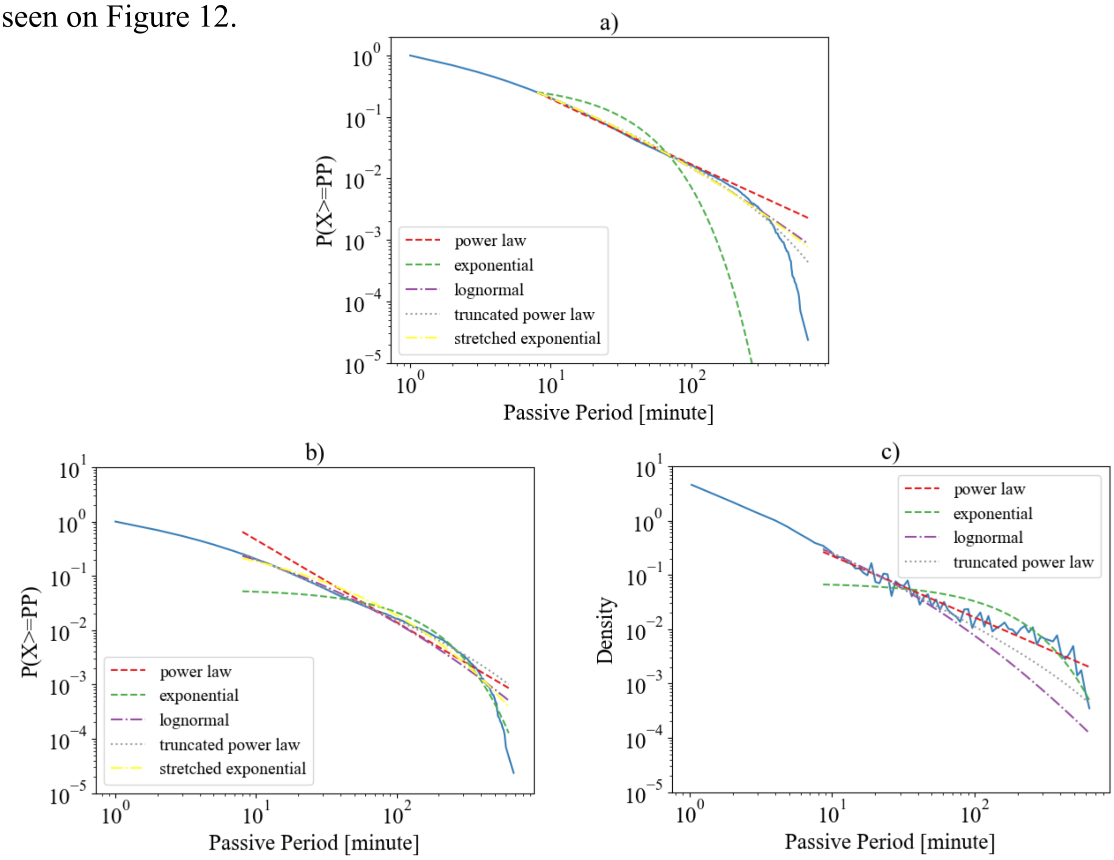

Figure 10. ***a**,The pooled, non-rescaled CCDF of passive periods, calculated using 100% of the mean as the threshold and using the ZCM(FMpost) activity determination method, with the different models fitted using MLE and xmin=8* ***b**, the same, but with the LS method* ***c**, the same, but on the PDF calculated with 100 bins and fitted with the LS method.*

The fitted models can be ranked based on their goodness of fit values. However, the different metrics can favour different properties of the fitted models. Figure 11 compares the best fit according to the goodness of fit metrics found in the reviewed studies. The LRT and the reduced chi-square metrics give approximately the same result, while the SSE, AIC and BIC metrics give different results from *xmin* = 6 or 7 minutes to *xmin* = 20 minutes. However, altogether the estimated best models seem to be somewhat consistent, lognormal when the whole range is fitted, and stretched exponential or exponential when only the tail is fitted, which undoubtedly is truncated in the case of the CCDF, as it was said in section 2.2.1.

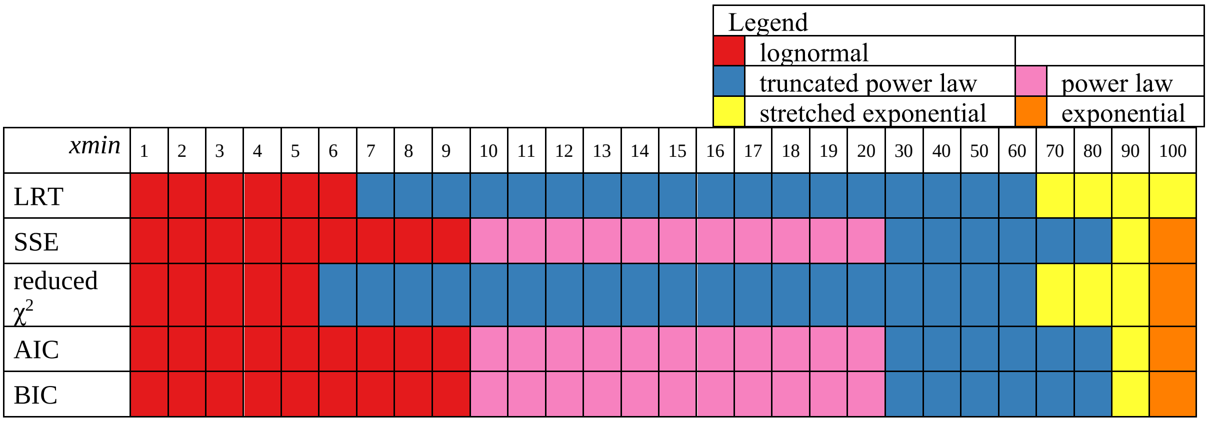

Figure 11 *The results of the estimation of the best models, calculated with different goodness of fit metrics as the different rows, and for different xmins (columns). The red cells represent lognormal, the blue cells truncated power law, the yellow cells stretched exponential, the pink cells power law and the orange cells exponential distributions.*

Altogether, both the estimation of the fitted models and choosing between them is affected by the different methods, therefore could lead to different results. However, sometimes only the power law is fitted to the passive periods, and only the lognormal to the active periods [12], in that case the goodness of fit metrics are not influencing the results.

4.7. Case study

To investigate the overall impact of the methodological differences in the statistical analysis of human activity I reproduced some of the different steps of several of the most important studies from the last 20 years using the same 10-day-long raw acceleration recordings of the research group. I attempted to use the same activity determination methods as the studies, however, in some cases the preprocessing method is only a guess or could not be determined at all, because it was not specified in the paper (for example, in the cases of [17] [9] [12]). The threshold rule of each study was also taken into account. My goal was to show that, while using the same dataset, the methodological differences heavily impact the estimated best fitting model.

On Figure 12 the CCDF and PDF of our data can be seen, calculated with the combination of the activity determination methods and threshold rules used in the studies. In Figure 13 each column represents a specific study's settings, and the different rows are different *xmins* used for the fitting range. Figure 13a contains the fitting of the passive periods, and Figure 13b contains the fitting of the active ones. The colours represent the model that the previously mentioned Python `powerlaw` package [21] determined as the best fit. It can be seen that just by changing one aspect of the method, various results could be obtained. Their shapes differ greatly already (without considering every methodological difference, e.g. pooling, rescaling, etc., thus their shapes becoming even more different), so it is no wonder that the model fitting (which also has different determining factors, as shown in the previous section (4.6)) gives varying results for them.

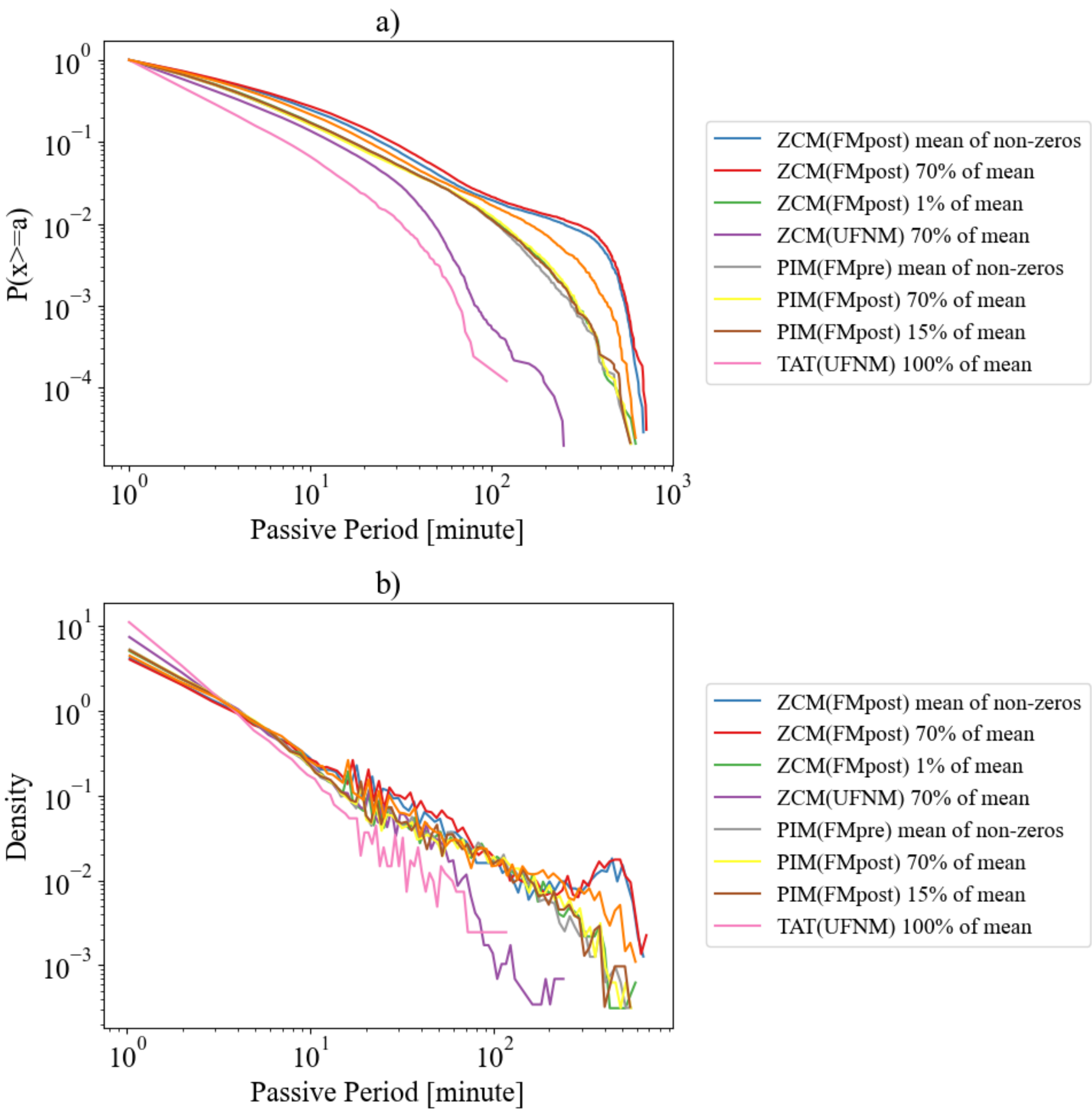

***Figure 12. a**, The pooled, non-rescaled CCDF of passive periods for the different types of activity determination methods and threshold values found in the articles **b**, the same for PDF*

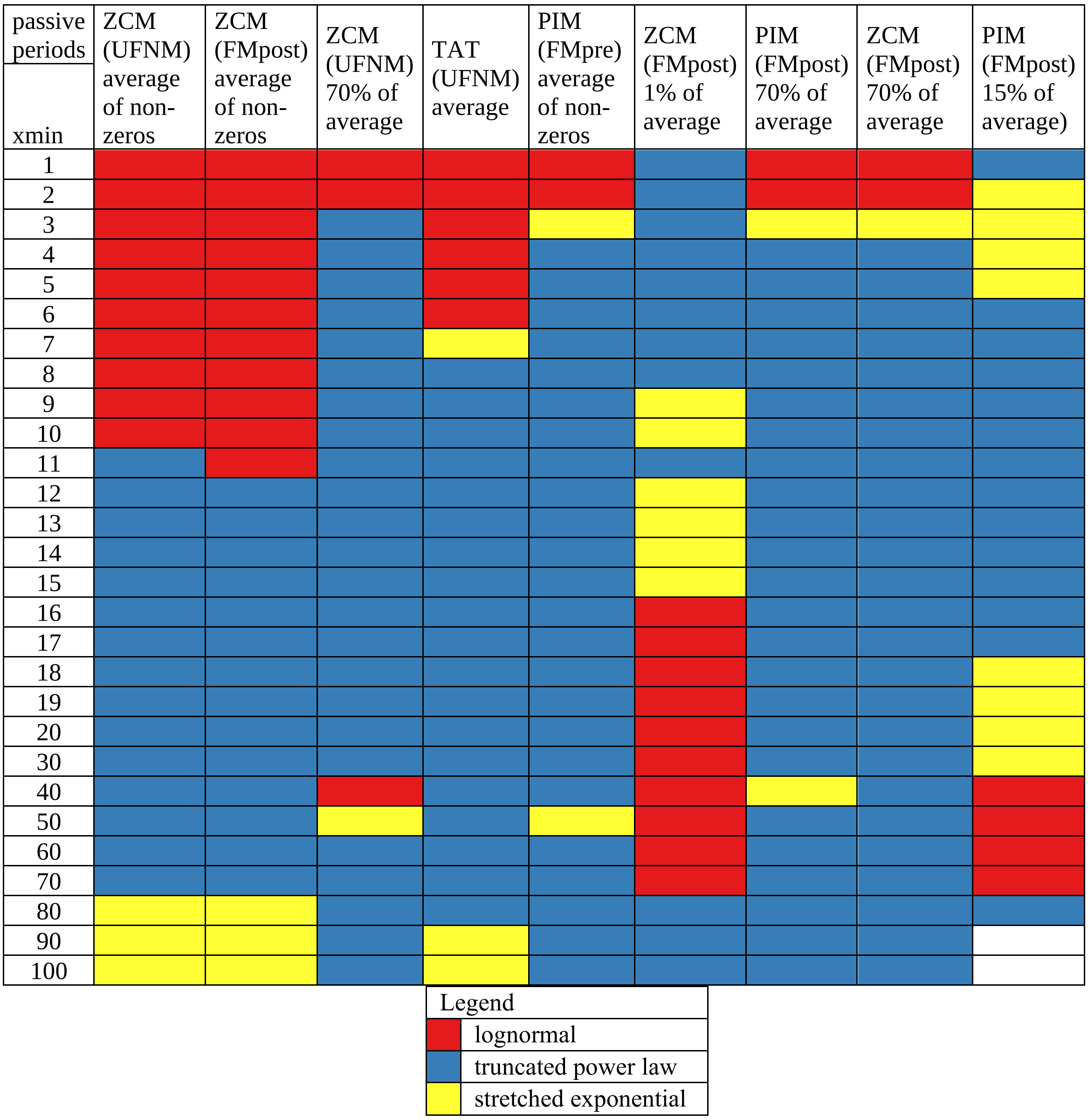

Figure 13. a, *The difference in the best estimated model for the different methods the articles used (columns) and different xmins (rows) in the case of the passive periods. The red cells represent lognormal, the blue cells truncated power law, and the yellow cells stretched exponential. In the case of white cells, a fit could not be determined because the distribution curve was too short and there was not enough data to determine it*

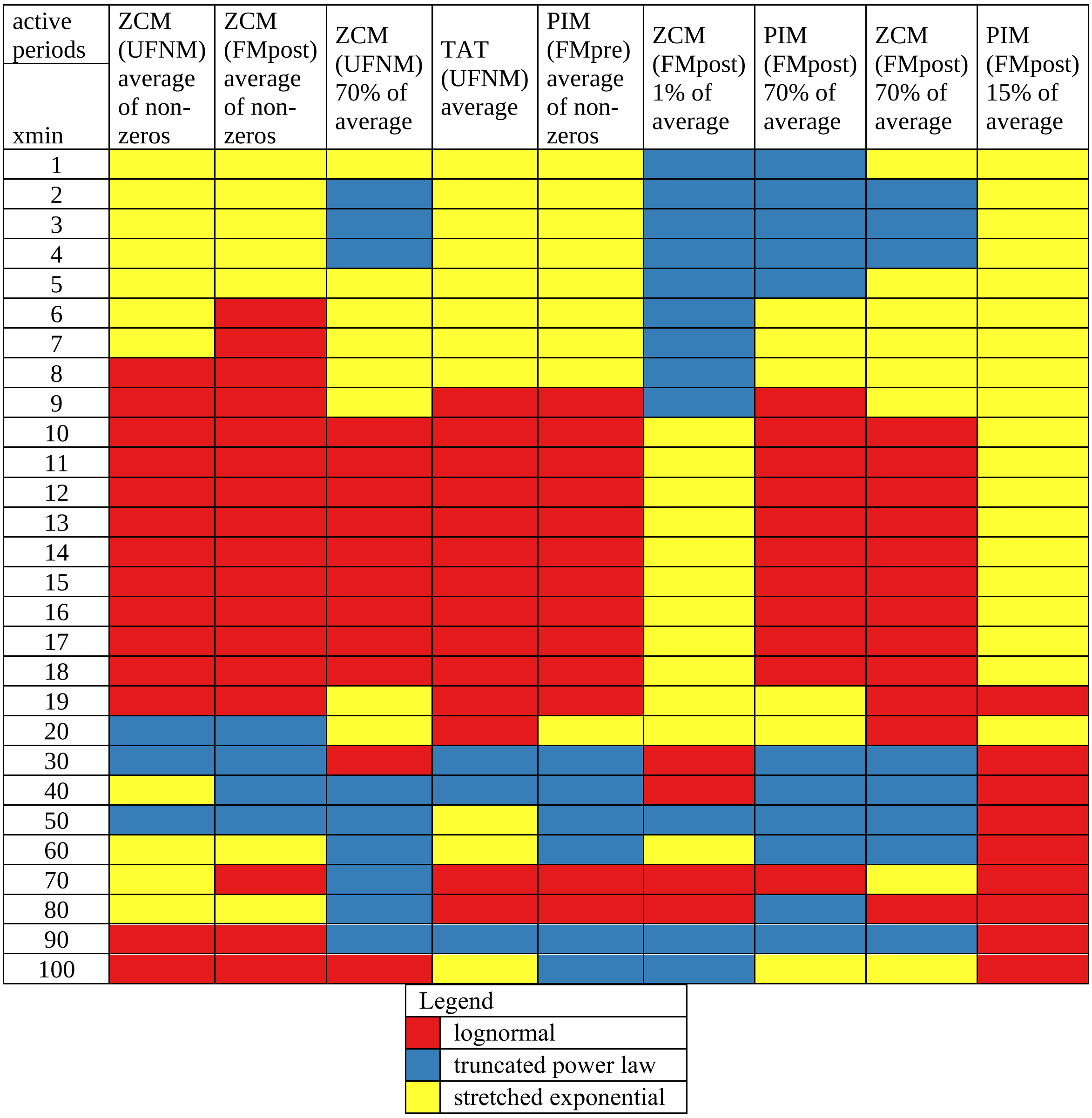

***Figure 13. b**, The difference in the best estimated model for the different methods the articles used (columns) and different xmins (rows) in the case of the active periods. The red cells represent lognormal, the blue cells truncated power law, and the yellow cells stretched exponential.*

In contrast, the research group showed previously, that the spectra of both acceleration signals, and activity values derived from them follow universal $1/f$ noise-based characteristics, as shown on Figure 14 which represents the spectral scale-free property, regardless of the actigraphic methodology and the threshold rule. The only variable, that can slightly affect the results is what the exact value of the corner frequency in the PSD is, and what could be the range of the fit of the $1/f^{\alpha}$ function, but it does not have a significant effect on the α of the power law distribution.

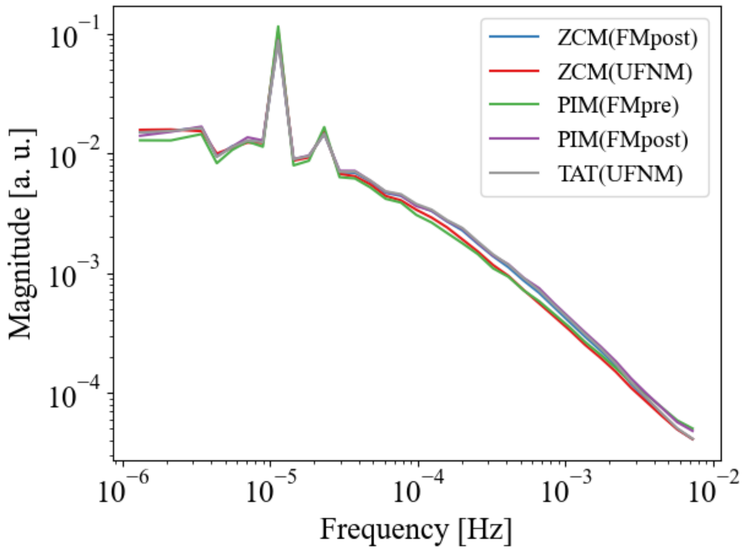

Figure 14. *The PSD of the activity signals used in the articles.*

5. Summary

Human dynamics exhibit scale-free characteristics in locomotor activity measurements. However, conflicting models have emerged from the statistical analyses of different datasets beyond the power law distribution. We presumed that it could be because of the non-standardized activity determination methods and statistical approaches.

Our objective was to find the underlying rules behind the contradicting results and show that the results of the statistical approach to the analysis of human locomotor activity is very dependent on the fitting methods, which are not the same in the different studies.

In the first part of the analysis of the activity values, we have shown that both the activity determination and the applied threshold rule affect the shape of the distribution, which we already published [2].

The next part of this study was to investigate the effect of the different statistical approaches, found in the relevant studies on the model-fitting. It was found that the bin number, the pooling, the method of fit and the used goodness of fit metrics all affect the results of the analysis. The threshold rule seems to affect the shape of the distribution, as well as the length of it. This result is consistent with what we have seen earlier, that the studies who determined the best fit as, or only fit a model other than power law or truncated power law used a threshold smaller than or equal to the mean of their dataset, because when a higher threshold is used in our case, the distribution seems more like the truncated power law. The *xmin* of the fit greatly affects the results, because when the fitting on the whole range, the predominantly obtained fit is the lognormal, while with increasing the *xmin*, thus fitting the tail of the curve more and more, first the truncated power law takes its place, and when only fitting the tail, the stretched exponential is the dominant fit. Ultimately, a case study was conducted, where we could see that these differences together greatly affect the results, while the spectra of the same activity values follow universal $1/f$ noise-based characteristics, which represents the spectral scale-free property.

This reinforces us that the scale-free nature of human activity can be examined more robustly in the frequency domain.

References

[1] B. Maczák, Z. Gingl, and G. Vadai, “General spectral characteristics of human activity and its inherent scale-free fluctuations,” *Scientific Reports*, vol. 14, no. 1, p. 2604, Jan. 2024, doi: https://doi.org/10.1038/s41598-024-52905-8.

[2] B. Maczák, K. P. Komáromi, and G. Vadai, “Different Aspects of the Scale-Free Nature of Human Activity – Examination of Its Spectral and Statistical Properties,” *2023 International Conference on Noise and Fluctuations (ICNF)*, pp. 1–4, Oct. 2023, doi: https://doi.org/10.1109/icnf57520.2023.10472755.

[3] A. M. Reynolds, “Scale-free animal movement patterns: Lévy walks outperform fractional Brownian motions and fractional Lévy motions in random search scenarios,” *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 43, p. 434006, Oct. 2009, doi: https://doi.org/10.1088/1751-8113/42/43/434006.

[4] M. C. González, C. A. Hidalgo, and A.-L. Barabási, “Understanding individual human mobility patterns,” *Nature*, vol. 453, no. 7196, pp. 779–782, Jun. 2008, doi: https://doi.org/10.1038/nature06958.

[5] L. Alessandretti, P. Sapiezynski, S. Lehmann, and A. Baronchelli, “Multi-scale spatio-temporal analysis of human mobility,” *PLOS ONE*, vol. 12, no. 2, p. e0171686, Feb. 2017, doi: https://doi.org/10.1371/journal.pone.0171686.

[6] B. Maczák, G. Vadai, A. Dér, I. Szendi, and Z. Gingl, “Detailed analysis and comparison of different activity metrics,” *PLOS ONE*, vol. 16, no. 12, p. e0261718, Dec. 2021, doi: https://doi.org/10.1371/journal.pone.0261718.

[7] J. G. Acker *et al.*, “The role of actigraphy in sleep medicine,” *Somnologie*, vol. 25, no. 2, pp. 89–98, May 2021, doi: https://doi.org/10.1007/s11818-021-00306-8.

[8] K. Krane-Gartiser, A. Asheim, O. B. Fasmer, G. Morken, A. E. Vaaler, and J. Scott, “Actigraphy as an objective intra-individual marker of activity patterns in acute-phase bipolar disorder: a case series,” *International Journal of Bipolar Disorders*, vol. 6, no. 1, Mar. 2018, doi: https://doi.org/10.1186/s40345-017-0115-3.

[9] M. Hirose *et al.*, “Altered distribution of resting periods of daily locomotor activity in patients with delayed sleep phase disorder,” *Frontiers in Psychiatry*, vol. 13, Oct. 2022, doi: https://doi.org/10.3389/fpsyt.2022.933690.

[10] S. E. Huber *et al.*, “Assessment of Fractal Characteristics of Locomotor Activity of Geriatric In-Patients With Alzheimer’s Dementia,” *Frontiers in Aging Neuroscience*, vol. 11, Oct. 2019, doi: https://doi.org/10.3389/fnagi.2019.00272.

[11] Ole Bernt Fasmer, E. Hauge, J. Ø. Berle, S. Dilsaver, and K. J. Oedegaard, “Distribution of Active and Resting Periods in the Motor Activity of Patients with Depression and Schizophrenia,” *Psychiatry Investigation*, vol. 13, no. 1, pp. 112–112, Jan. 2016, doi: https://doi.org/10.4306/pi.2016.13.1.112.

[12] András Búzás *et al.*, “Hierarchical organization of human physical activity,” *Scientific Reports*, vol. 14, no. 1, Mar. 2024, doi: https://doi.org/10.1038/s41598-024-56185-0.

[13] D. Chialvo *et al.*, “How we move is universal: Scaling in the average shape of human activity,” *Papers in Physics*, vol. 7, p. 070017, Nov. 2015, doi: https://doi.org/10.4279/pip.070017.

[14] T. Nakamura *et al.*, “Of Mice and Men — Universality and Breakdown of Behavioral Organization,” vol. 3, no. 4, pp. e2050–e2050, Apr. 2008, doi: https://doi.org/10.1371/journal.pone.0002050.

[15] J. J. Chapman, J. A. Roberts, V. T. Nguyen, and M. Breakspear, “Quantification of free-living activity patterns using accelerometry in adults with mental illness,” *Scientific Reports*, vol. 7, no. 1, Mar. 2017, doi: https://doi.org/10.1038/srep43174.

[16] J. K. Ochab *et al.*, “Scale-Free Fluctuations in Behavioral Performance: Delineating Changes in Spontaneous Behavior of Humans with Induced Sleep Deficiency,” *PLoS ONE*, vol. 9, no. 9, pp. e107542–e107542, Sep. 2014, doi: https://doi.org/10.1371/journal.pone.0107542.

[17] M. Kawabata *et al.*, “Temporal organization of rest defined by actigraphy data in healthy and childhood chronic fatigue syndrome children,” *BMC Psychiatry*, vol. 13, no. 1, Nov. 2013, doi: https://doi.org/10.1186/1471-244x-13-281.

[18] J.-H. Lee *et al.*, “Statistical properties of human activity and criticality in active behavior,” *EPL (Europhysics Letters)*, vol. 126, no. 6, p. 68001, Jul. 2019, doi: https://doi.org/10.1209/0295-5075/126/68001.

[19] N. Marshall, N. M. Timme, N. Bennett, M. Ripp, E. Lautzenhiser, and J. M. Beggs, “Analysis of Power Laws, Shape Collapses, and Neural Complexity: New Techniques and MATLAB Support via the NCC Toolbox,” *Frontiers in Physiology*, vol. 7, Jun. 2016, doi: https://doi.org/10.3389/fphys.2016.00250.

[20] S. M. Burroughs and S. F. Tebbens, “Upper-truncated Power Laws in Natural Systems,” *Pure and Applied Geophysics*, vol. 158, no. 4, pp. 741–757, Apr. 2001, doi: https://doi.org/10.1007/pl00001202.

[21] J. Alstott, E. Bullmore, and D. Plenz, “powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions,” *PLoS ONE*, vol. 9, no. 1, p. e85777, Jan. 2014, doi: https://doi.org/10.1371/journal.pone.0085777.

[22] J. Bai *et al.*, “An Activity Index for Raw Accelerometry Data and Its Comparison with Other Activity Metrics,” *PLOS ONE*, vol. 11, no. 8, p. e0160644, Aug. 2016, doi: https://doi.org/10.1371/journal.pone.0160644.

[23] Balint Maczak, A. Antal, and Gergely Vadai, “The Noise of Our Daily Motion – General Spectral Characteristics of Human Mobility and Activity,” *Fluctuation and Noise Letters*, Oct. 2024, doi: https://doi.org/10.1142/s0219477524400613.