Exploratory Graphics for Functional Actigraphy Data

Jürgen Symanzik*   William Shannon†

Abstract

Actigraphy is an emerging technology for measuring a patient's overall activity level nearly continuously over time. An actigraph is a watch–like device attached to the wrist or ankle that uses an accelerometer to measure movement every minute, or even more often. Actigraphy data can best be described as functional data. In this paper, we will present exploratory graphical displays for various components of functional actigraphy data.

Key Words: Actigraphy, Functional Data Analysis, Visualization.

1. Introduction

Actigraphy is an emerging technology for measuring a patient's overall activity level nearly continuously over time. An actigraph (see Figure 1) is a watch–like device attached to the wrist or ankle that uses an accelerometer to measure movement every minute, or even more often. An *American Academy of Sleep Medicine* (AASM) report defines the practice parameters for the use of actigraphy in the study of sleep and circadian rhythms (Ancoli-Israel et al. 2003, Morgenthaler et al. 2007). The AASM report recommends, among other things, that actigraphy can be a useful tool for detecting sleep in healthy individuals and a useful adjunct to a detailed history and subjective sleep diary for diagnosing and treating various sleep disorders. In addition, an actigraph can be used continuously over several days because of its small size and portability, so it has the capability of obtaining data that polysomnography alone cannot provide. According to the AASM report, there is also a growing interest in actigraphy as a tool for objectively measuring fatigue.

To increase the use of actigraphy as a tool for objectively measuring fatigue, additional robust data models and statistical data analysis methods and software are needed. Current analysis tools have focused on sleep assessment and use graphical displays with zoom in/out capabilities and basic summary statistics (e.g., mean and standard deviation of the activity level within a patient) that enable the clinician to view a single patient's actigraphy data to identify aberrant patterns of activity. The clinician then relates these aberrant patterns to clinical assessment of the patient, a subjective and time–consuming process. Group comparisons reduce multivariate actigraphy data to summary statistics (e.g., mean, standard deviation) which throws out important information such as the pattern of activities. Furthermore, actigraphy data show considerable variability within and between patients, so that "landmark" analyses may not be powerful enough to detect changes in activity levels due to disease or treatment response.

To overcome these data analysis limitations, three specific aims have been proposed that will be implemented over the next few years:

1. We will define an object oriented data model for actigraphy and patient–level data. This data model will use object oriented programming strategies programmed in the R open source statistical software programming language to store patient–level (e.g., diagnoses, demographics, clinical variables, treatments, medications, previous history, research information) and actigraphy–level data in a relational database. Help menus, input/output functions, basic summaries and printing tools, data transformations, and statistical, visualization, and functional data analysis (FDA) algorithms developed in Aim 2 will be programmed in simple pull down menus.

2. We will apply existing and develop new advanced statistical and visualization methods for actigraphy data and include them in the data model defined in Aim 1. These will be developed to answer clinically relevant questions (e.g., do activity profile patterns correlate with disease severity?).

3. To validate and improve the tools developed in Aims 1 and 2, we will prospectively collect actigraphy data under controlled experimental conditions (for calibration and methodology development) and in several patient subgroups.

---

*Utah State University, Department of Mathematics and Statistics, 3900 Old Main Hill, Logan, UT 84322–3900, USA. Phone: 435 797 0696, Fax: 435 797 1822, E-mail: `symanzik@math.usu.edu`

†Washington University School of Medicine, Department of Medicine and Division of Biostatistics, 660 S. Euclid Ave., Campus Box 8005, St. Louis, MO 63110, USA. Phone: 314 454 8356, Fax: 314 454 5113, E-mail: `wshannon@wustl.edu`

Figure 1: (a) What looks like a regular watch from the distance turns out to be an actigraph (b), here an Actiwatch from Respironics, Inc. (c) Via an ActiReader device from Respironics, Inc., data is transferred every few weeks from the Actiwatch to a computer where it is stored and (d) displayed on the computer screen.

Within this paper, we will concentrate on the visualization component listed in Aim 2. We will discuss existing visualization techniques for actigraphy data in Section 2. In Section 3, we provide a brief summary of existing visualization techniques for functional data. In Section 4, we will present some of our suggested new visualization approaches to actigraphy data and demonstrate how these can reveal new insights into actigraphy data. We will finish with a discussion in Section 5.

2. Current Visualization of Actigraphy Data

A typical actigraph data set on a patient is presented in Figure 2(a). Features of this display include the activity level indicated by the black spikes measured at a specific time of the day (x–axis) over multiple days (sequence of horizontal lines ranging from April 26, 2007, through May 7, 2007). This display is produced by the actigraph manufacturer's software and represents a state–of–the–art display and analysis tool of actigraphy data. From this plot, it is clear that there is significant minute–to–minute variability within each day and significant day–to–day variability across the rows. For example, overall activity levels are much higher on May 6, 2007, during daytime than on most other days. Low activity at night is expected and found as indicated by the low activity levels recorded between 2:00 am and 9:00 am. The red line segments indicate where the patient most likely is awake or makes major movements while asleep. The blue triangles indicate time points marked by the patient where a new activity started, such as going to bed or getting out of bed, driving to work, arriving at work, etc.

Figure 2(b) displays histograms of the activity levels of four patients. In this display, the x–axis indicates the level of activity (higher represents more active) and the y–axis shows the number of times the patient exhibits that level of activity. The higher the histogram bar, the more times the patient had the activity level indicated on the x–axis. The top two patient histograms show that most activity levels are very low (high bars on the left side of the histogram and no bars on the right). The lower two patient histograms show periods of high activity as represented by the histogram bars that stretch much further to the right. This interpretation is confirmed by summary statistics of mean activity levels of 5.4, 20.4, 209.7, and 206.5, for patients C, D, A, and B, respectively.
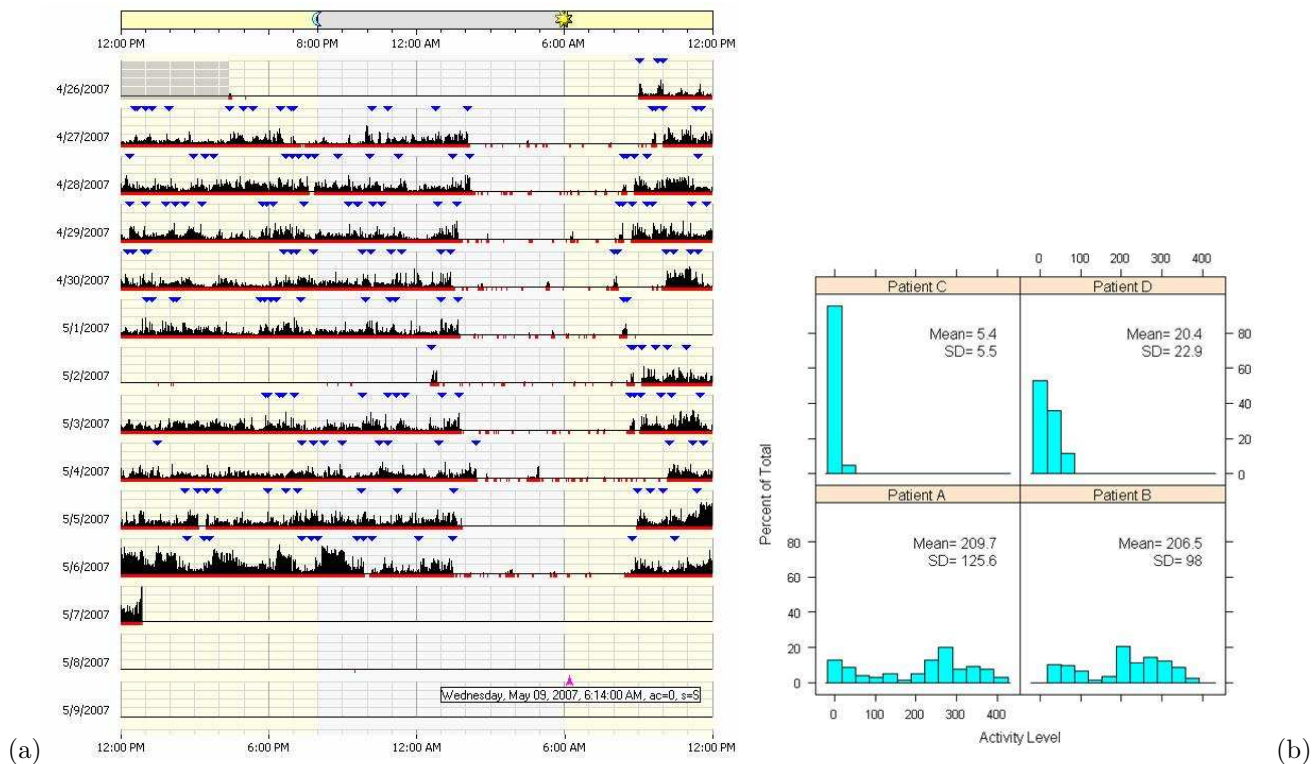
Figure 2: (a) Default state–of–the–art display of actigraphy data for a single patient, as provided by the actigraph manufacturer's software. (b) Numerical and graphical summaries of actigraphy data of four different patients.

3. Current Visualization of Functional Data

Ramsay & Silverman (2006), page 38, characterize functional data as follows: *"The basic philosophy of functional data analysis is to think of observed data functions as single entities, rather than merely as a sequence of individual observations."* Therefore, visualization techniques and exploratory graphics for functional data should be applicable to functions of data, and not just to single observations.

Unfortunately, existing tools for visually exploring functional data are rather limited. Throughout their book, Ramsay & Silverman (2006) use standard displays for time series to display functional data. In addition, they use so–called phase plane plots which show the second derivative (i.e., acceleration) of the function on the vertical axis and the first derivative (i.e., velocity) on the horizontal axis. Derivatives play an important role in the analysis of functional data. In Section 8.3, Ramsay & Silverman (2006) point out the importance of visualizing the results of functional principal components analysis (PCA). The R package *fda* provides implementations for most of the methods discussed in Ramsay & Silverman (2006) and the accompanying Web page `http://www.functionaldata.org` provides direct access to Matlab, R, and S–Plus code.

Ferraty & Vieu (2006) use about the same graphical displays as Ramsay & Silverman (2006) throughout their book. In addition, Ferraty & Vieu (2006) work with plots of probability curves and estimated densities that are based on these probability curves, as discussed in Section 9.4 of their book. Related R and S–Plus routines can be obtained from the accompanying Web page `http://www.lsp.ups-tlse.fr/staph/npfda/`.

Advanced interactive software aimed at the visualization of functional data is rare. Jank et al. (2008), page 887, indicate that *"it is more challenging to visualize functional data than classical data. The visualization process is often further complicated by the coupling of functional observations with cross–sectional attribute data. [. . .] All of these characteristics correspond to cross–sectional information in that they do not change [. . .]. The coupling of time series with cross–sectional information is important because the relationship between the two could be the main aim or at least one of the aims of the analysis."* In a medical scenario, some possible cross–sectional attributes that relate to human activity levels could be gender, race, age, season (winter versus summer), and time of data collection (baseline, during treatment, or after treatment). Jank et al. (2008) describe *TimeSearcher*, a rare example of a visualization tool that provides an interactive approach to the visual exploration of functional data and time series data. An extended version of *TimeSearcher* that specializes on the analysis of online auction data, combined with additional attribute data, can be obtained from the Web page `http://www.cs.umd.edu/hcil/timesearcher/`.
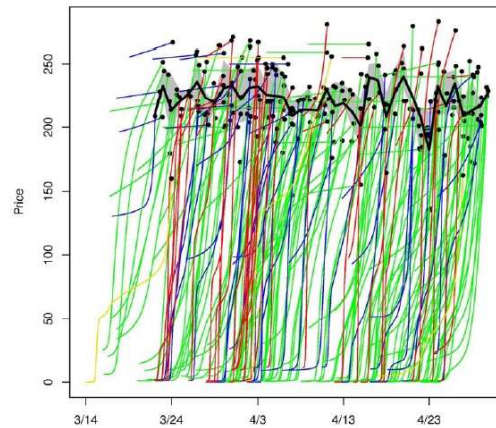
Figure 3: Rug plot displaying the price evolution of more than 200 online auctions over time. The x–axis shows time, the y–axis indicates the development of the sales price over time. The black dot indicates the final sales price of an auction. Different colors are used to distinguish among different auction lengths. The black line indicates the daily average final price, and the gray band indicates the daily interquartile range. This figure has been taken from `http://www.smith.umd.edu/faculty/wjank/DIV-Berlin2006.pdf`, page 30, and has been reprinted with permission of the author.

Figure 3 shows a rug plot that displays the price evolution of more than 200 online auctions over time. An apparent limitation of this plot is the difficulty to distinguish which of the auctions are similar. Does the auction length (colored as yellow for three days, blue for five days, green for seven days, and red for ten days) have an impact on the price evolution?

## 4. Future Visualization of Actigraphy Data

Histograms and summary statistics give a good general description of how the patients differ in activity level, but they fail to capture any detail of the time course of activity patterns. Figure 4 present the same four patient's activity levels (previously shown in Figure 2(b)) as functional data over the day. In these graphs, the x–axis shows the time of the day (midnight on the left to midnight the following night on the right) and the y–axis shows the activity level. The raw activity levels have been replaced by smoothed functions (the solid and dashed lines fit by loess regression with two different parameter settings) to show the activity pattern without the random noise inherent in any data. In other words, we have replaced the scatter plot display of activity level by time with a smoothed function of activity level by time. These activity patterns correspond to the histograms in Figure 2(b) with little activity anytime during the day for the top two patients (C and D) and "normal" activity patterns for the lower two patients (A and B). By "normal", we mean low activity during sleeping hours, rising activity in the morning, relatively constant and high activity throughout the day, and decreasing activity throughout the evening. More discussion of these activity pattern profiles will be given later in this section.

While the display of smoothed continuous curves for the daily activity pattern as shown in Figure 4 is a first step towards a modern statistical visualization of these data, additional visualization features are desirable, in particular when taking into account that actigraphy data can be considered as functional data.

In the remainder of this section, we will present statistical displays for several variations of the collected actigraphy data. In particular, these displays should be useful when comparing data for a single patient over time (e.g., several days at baseline versus several days after treatment) or by comparing different patient groups (patients from the treatment group versus patients from the control group). Moreover, a clinician working with these displays should be able to focus on some of the previously mentioned cross–sectional attributes that relate to human activity levels, such as gender, race, age, season (winter versus summer), and time of data collection (baseline, during treatment, or after treatment). Advanced graphical displays for actigraphy data should build on current displays suitable for functional data, as introduced in Section 3, and these displays should allow to interact with multiple views via brushing and linking (see, for example, Symanzik 2004, for a defintion of these terms). It is desirable to display the following types of actigraphy–related data:
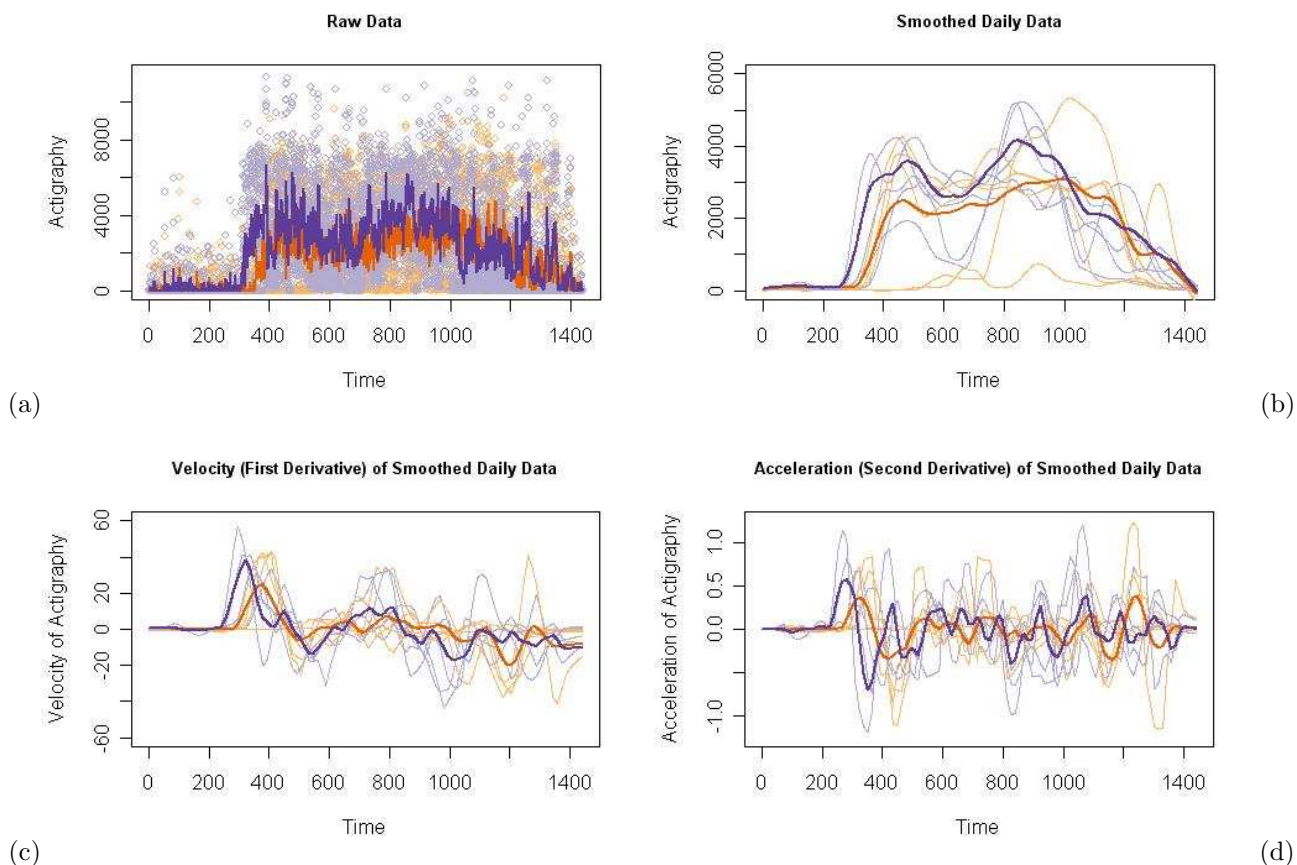
- Raw data

- Smoothed data
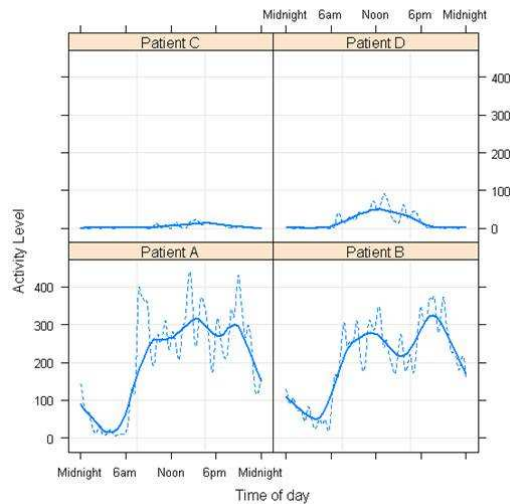
Figure 4: Activity levels of the same four patients as shown in Figure 2(b). The x–axis shows the time of the day (from midnight to midnight) and the y–axis shows the activity level. The raw activity data have been smoothed by two different loess smoothers and are displayed as continuous curves.

- Averages and other summaries

- Velocity (First Derivative)

- Acceleration (Second Derivative)

- Cumulative Sums

- Sorted Cumulative Sums

In our next example, we follow one single patient and compare activity levels for five days at baseline (orange) with activity levels for five days after treatment with a cognitive behavior therapy (CBT) after six months (purple) and further describe the advanced displays introduced in the previous paragraph. These data originate from a pilot study conducted by Kenneth E. Freedland at the Washington University of St. Louis, School of Medicine (private communication). In this pilot study, 23 depressed patients with heart failure were enrolled. Patients were randomly assigned to six months of cognitive behavioral therapy or usual care (UC). Patients receiving CBT showed statistically significant improvement over UC subjects in several mental health, quality of life, and cardiac functioning measures. In addition, for eleven of the patients, actigraphy data were collected at baseline and at the six–months follow–up. Here, we just concentrate on a single patient, but comparing the actigraphy data for the CBT group and the UC group also would have been of interest. In each of the next three figures, the horizontal axis represents time in minutes and ranges from 0 minutes to $24 \times 60 = 1440$ minutes. Thus, actigraphy–related data for each of the ten days (five at baseline, five at follow–up) start at 0 minutes.

Figure 5(a) shows the raw activity level for this single patient for five days of actigraphy measured at baseline (light orange dots) and for five days measured at the six–months follow–up (light purple dots). Dark orange dots indicate the daily average activity level for each minute at baseline, while dark purple dots indicate the daily average activity level at follow–up. Due to the extensive overplotting of points, it is difficult to identify a clear pattern, but it appears that the average activity level is increased through most times of the day at the follow–up period. Figure 5(b) overcomes this overplotting issue by showing lowess–smoothed curves (with parameter $f = 0.1$) for each of the five days at baseline (light orange curves), the baseline average (dark orange curve), the five days at follow–up (light purple curves), and the follow–up average (dark purple curve). The hypothesis that the average activity level is increased at follow–up becomes more evident. In particular, there are two days at baseline that show very little activity up to 750 minutes into the day (which is equivalent to 12:30 hours). Figure 5(c) shows velocity, i.e., the first derivative, of the smoothed daily activity data, and Figure 5(d) shows acceleration, i.e., the second derivative, of the smoothed daily activity data. Velocity tells us how the activity level is changing instantly over time. Is the patient becoming more active, less active, or staying at the same activity level? Acceleration tells us how quickly these changes in velocity are occurring. Does the patient have a sudden spurt in energy so the patient becomes more active more quickly versus a patient who becomes more active very slowly and appears to be struggling. Velocity
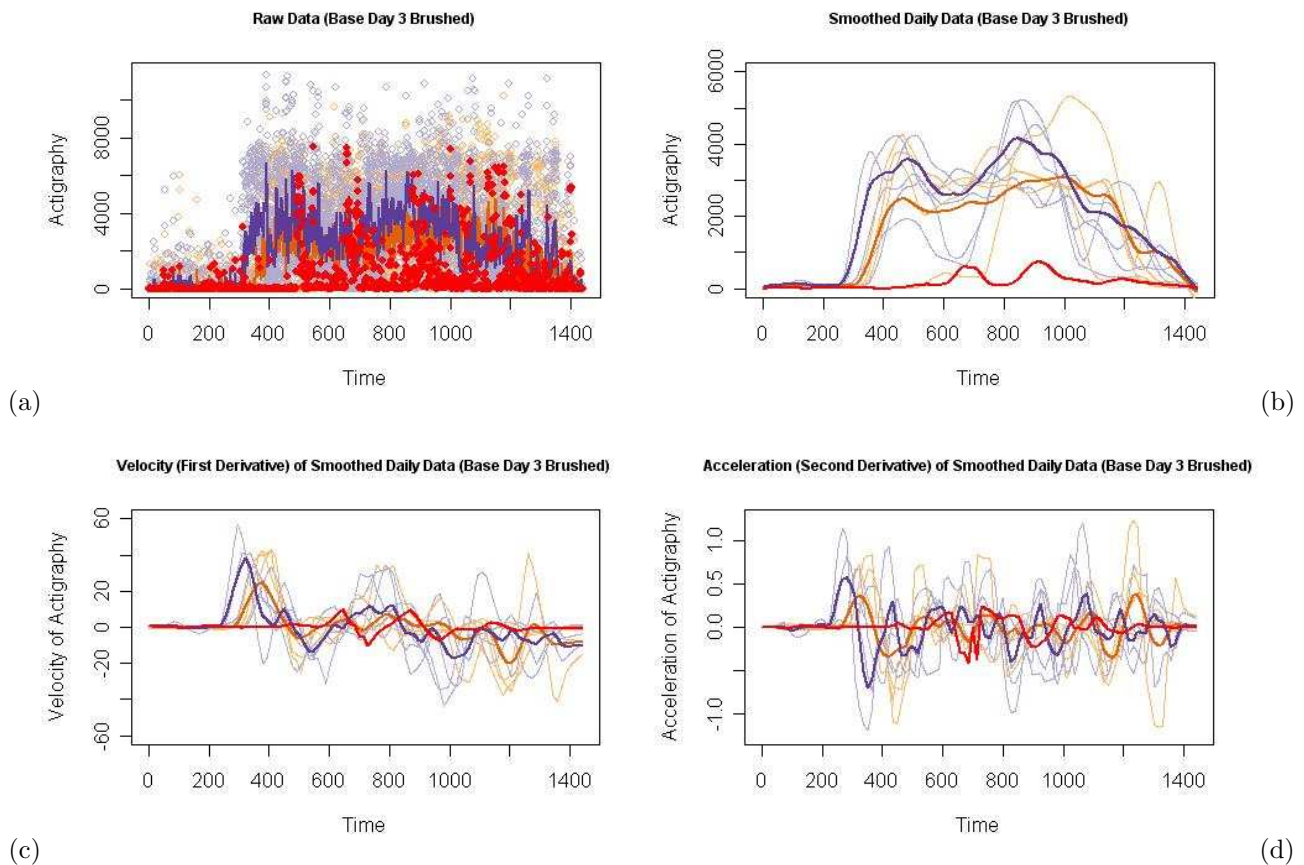
Figure 5: (a) Raw data, (b) smoothed data, (c) velocity of smoothed data, and (d) acceleration of smoothed data for five days at baseline and five days at follow–up. Light orange points and curves represent data from a single day at baseline, light purple points and curves represent data from a single day at follow–up, dark orange points and curves represent the five–day average at baseline, and dark purple points and curves represent the five–day average at follow–up. The horizontal axis represents time in minutes, ranging from 0 minutes to 1440 minutes.

and acceleration functions may allow us to distinguish subjects who respond quickly or slowly to their daily activities and indicate how fatigue and clinical measures impact behavior. The same color coding has been used for these two plots as for the other plots in this figure. It is apparent that the average velocity at follow–up is much higher than at baseline and starts earlier in the day. Also, acceleration, i.e., the change of velocity, is higher at follow–up than at baseline. Overall, the dark orange curve could be interpreted as the "normal" activity pattern at baseline while the dark purple curve could be interpreted as the "normal" activity pattern at follow–up. Low levels at night when the patient is sleeping, increasing in the morning when the patient wakes up, relatively high and constant throughout the day, and decreasing in the evening.

As observed before, there are two days at baseline that show very little activity throughout the day. Figure 6 shows exactly the same plots as Figure 5, with the exception that the day with the lowest overall activity at baseline (Day 3) has been brushed in red. This means, the orange points and curves for Day 3 from Figure 5 have been replaced with red points and curves here. All plots in this figure reveal that this single day is truly an outlier with some rather unusual activity patterns throughout the day.

An immediate question is to learn more details about the second day at baseline that shows very little activity throughout the day. Figure 7 deals with this particular day at baseline (Day 2) that has been brushed in gray throughout this figure. Figure 7(a) shows that during Day 2, between a time of around 900 minutes to 1200 minutes (15:00 hours to 20:00 hours), the patient had the highest activity level during this five–hour period this particular day when compared with all other days (at baseline and at follow–up). This leads to the question whether the overall activity was indeed lower during Day 2, compared to other days, or whether main activities were mostly shifted into the afternoon and early evening hours.

This question can best be answered via the cumulative sum plot in Figure 7(b). This plot shows accumulated activity obtained by adding up the activity counts as one moves across the time (x) axis from midnight (far left) to
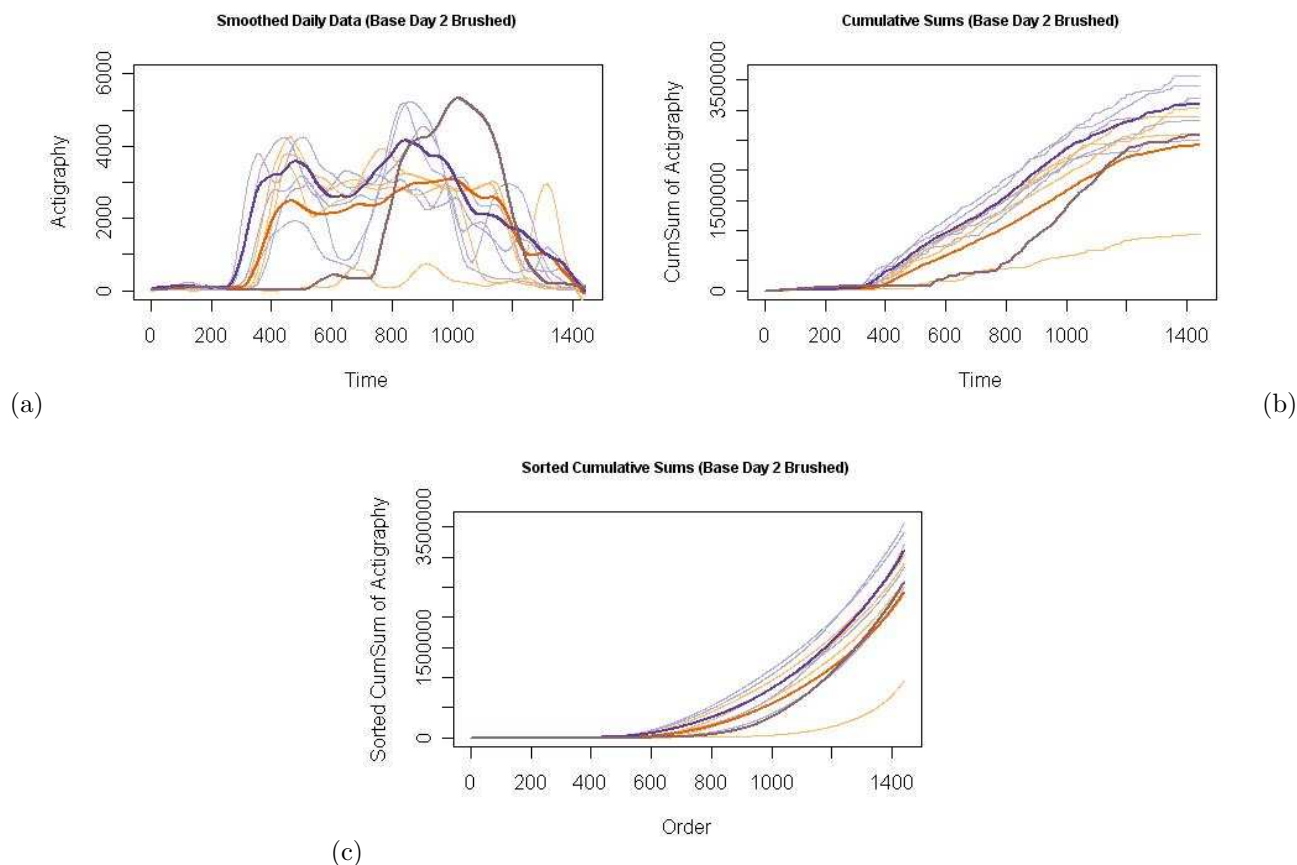
Figure 6: (a) Raw data, (b) smoothed data, (c) velocity of smoothed data, and (d) acceleration of smoothed data for five days at baseline and five days at follow–up. The same colors have been used as in Figure 5. In addition, red has been used to brush Day 3 at baseline.

midnight 24 hours later (far right). This plot is useful to show total activity of a patient up to a particular time of the day. Indeed, up to a time of 750 minutes, the patient had accumulated very little activity this day. However, from there on, the accumulation is steeper than for most other days and accumulation continues at a noticeable rate in the evening where most other curves are flat, indicating little activity. At midnight of Day 2, the total accumulated activity was comparable with other days at baseline (and even above the average at baseline that gets distorted due to the overall low activity pattern observed at Day 3 at baseline).

Finally, Figure 7(c) shows sorted accumulated activity obtained by adding up the activity counts from smallest to largest. It should be noted that the x–axis no longer represents time but order. Statistically, we are summing up the observed order statistics $x_{(1)}$ to $x_{(1440)}$ to create this plot. As can be seen from the dark orange line, a "normal" day at baseline has about 600 units (10 hours) of almost no activity and a "normal" day at follow–up has about 480 units (8 hours) of almost no activity, which easily could be interpreted as sleep. Day 2 at baseline has about 720 units (12 hours) of almost no activity, but then accumulates activity at a much higher rate than the other days at baseline. It should be noted that the total accumulated activity shown for each day in Figures 7(b) and 7(c) is the same; what differs is the order of accumulation. A clinical question of interest now is whether this extensive high activity seen in the afternoon and early evening of Day 2 at baseline has resulted in such an exhaustion that prevented this patient from almost any activity at Day 3 at baseline.

5. Discussion

In this paper, we have demonstrated how exploratory statistical graphics can be used for the visualization and interpretation of functional actigraphy data. Useful visualization techniques include plotting of raw data, smoothed data, and averages. Linking multiple plots and brushing in these plots to highlight individual days or individual patients is of particular interest. In addition to obvious plots, new ideas of plotting actigraphy data have been presented, e.g., the cumulative sums in Figure 7(b) and the sorted cumulative sums in Figure 7(c). Both of these

Figure 7: (a) Raw data, (b) cumulative sums, and (c) sorted cumulative sums of smoothed data for five days at baseline and five days at follow–up. The same colors have been used as in Figure 5. In addition, gray has been used to brush Day 2 at baseline.

representations present new perspectives for viewing actigraphy data that will be further studied to see if they provide additional clinical insight. In our ongoing work, we will explore other visualization methods as well to see what other clinical interpretations might be available for various statistical plots.

In summary, visualization of actigraphy data provides a potential for application in various medical fields. Moreover, using modern visualization techniques, additional insights into actigraphy data can be obtained that are currently unavailable when using traditional methods that are supported by actigraph manufacturer's software packages. Overall, these modern visualization techniques provide an ease to compare baseline and follow–up actigraphy data. This can be data for a single patient observed over time, or data for multiple patients (cross–sectional or longitudinal). Finally, these modern visualization techniques help to identify outliers (unusual days or unusual patients) as well as to compare averages, i.e., "normal" activity patterns for different time periods or different patient groups.

REFERENCES

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W. & Pollak, C. P. (2003), 'The Role of Actigraphy in the Study of Sleep and Circadian Rhythms', *Sleep* **26**(3), 342–392.

Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis — Theory and Practice*, Springer, New York, NY.

Jank, W., Shmueli, G., Plaisant, C. & Shneiderman, B. (2008), Visualizing Functional Data with an Application to eBay's Online Auctions, *in* C. Chen, W. Härdle & A. Unwin, eds, 'Handbook of Data Visualization', Springer, Berlin, Heidelberg, pp. 873–898 & 2 Color Plates.

Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson, A., Coleman, J., Lee-Chiong, T., Pancer, J. & Swick, T. J. (2007), 'Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007', *Sleep* **30**(4), 519–529.

Ramsay, J. O. & Silverman, B. W. (2006), *Functional Data Analysis (Second Edition)*, Springer, New York, NY.

Symanzik, J. (2004), Interactive and Dynamic Graphics, *in* J. E. Gentle, W. Härdle & Y. Mori, eds, 'Handbook of Computational Statistics — Concepts and Methods', Springer, Berlin, Heidelberg, pp. 293–336.