Articles in PresS. J Appl Physiol (June 25, 2015). doi:10.1152/japplphysiol.00026.2015

Title: Methods to estimate aspects of physical activity and sedentary behavior from high frequency wrist accelerometer measurements

Authors:

John Staudenmayer[1]

Shai He[1]

Amanda Hickey[2]

Jeffer Sasaki[2]

Patty Freedson[2]

Affiliations:

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003 USA.

[2]Department of Kinesiology, University of Massachusetts Amherst, Amherst, MA 01003 USA.

Address for Correspondence:

John Staudenmayer

Department of Mathematics and Statistics

University of Massachusetts Amherst

Amherst, MA 01003 USA

413 545 0999

jstauden@math.umass.edu

Running Head: Methods to process wrist accelerometer data

Abstract:

This investigation developed models to estimate aspects of physical activity and sedentary behavior from three-axis high frequency wrist worn accelerometer data. The models were developed and tested on twenty participants (n=10 males, n=10 females, mean age= 24.1, mean BMI = 23.9) who wore an ActiGraph GT3X+ accelerometer on their dominant wrist and an ActiGraph GT3X on the hip while performing a variety of scripted activities. Energy expenditure was concurrently measured by a portable indirect calorimetry system. Those calibration data were then used to develop and assess both machine learning and simpler models with fewer unknown parameters (linear regression and decision trees) to estimate METs and to classify activity intensity, sedentary time, and locomotion time. The wrist models, applied to 15-second windows, estimated METs (random forest: rMSE = 1.21 METs, hip: rMSE = 1.67 METs) and activity intensity (random forest: 75% correct, hip: 60% correct) better than a previously developed model that used counts per minute measured at the hip. In a separate set of comparisons, the simpler decision trees classified activity intensity (random forest: 75% correct, tree: 74% correct), sedentary time (random forest: 96% correct, decision tree: 97% correct), and locomotion time (random forest: 99% correct, decision tree: 96% correct), nearly as well or better than the machine learning approaches. Preliminary investigation of the models' performance on two free-living people suggests that they may work well outside of controlled conditions.

Keywords: ActiGraph, GT3X+, high-frequency, triaxial

Introduction

Processing accelerometer data to estimate aspects of physical activity and sedentary behavior remains a challenge. The original approach used data collected in the lab to obtain hip activity monitor counts and energy expenditure from a few or several activities (12),(9),(3),(21). These lab calibration data were used to build simple linear regression models for estimating energy expenditure or activity intensity from counts averaged over some user specified time frame (e.g. 15 sec, 1 min). The most widely used simple linear regression model for the ActiGraph single axis accelerometer worn on the hip was that developed by Freedson et al (9) and uses counts per minute as the single independent variable. Solving the regression for activity counts using the lower and upper boundaries of absolute moderate intensity activity as the dependent variable (3 and 6 METs) became known as the Freedson cut-point method. Despite its limitations, this model has endured multiple challenges and remains a popular method of choice for processing field based ActiGraph accelerometer data. A modified version of the original Freedson cut-points defining the count range for light, moderate and vigorous activity was used in the NHANES ActiGraph data analysis which is the largest nationally representative database for objectively monitored physical activity (23).

All of the ‘cut-point’ algorithms for adults that use derived activity counts as inputs into simple linear regression models are for accelerometers worn on the hip. Crouter et al (5) have developed wrist ‘cut-points’ for youths. One modification to the hip accelerometer activity count linear regression modeling was developed by Crouter et al (4, 6) where activity counts were directed to one of two regressions models based on the variability in the activity counts. That approach used the size of the standard deviation of activity counts to direct the activity counts to one of three equations: 1) resting level of energy expenditure; 2) walking and running; and 3) lifestyle activities. Another approach to processing accelerometer data is to use more sophisticated statistical methods, also known as statistical or machine learning techniques. These methods extend ‘cut-point’ algorithms in two ways. First, they summarize accelerometer signals using more than just the total count in an interval. Instead, they use statistical summaries of counts that describe both the distribution of acceleration in an interval and the temporal dynamics of the signal. After that, they use those statistical summaries as inputs to (or covariates for) models that estimate either energy expenditure or classify an aspect of the activity that is being performed. Examples of this general approach include (20),(25),(19),(17). In recent work, Lyden et al (15) developed and validated count-based machine learning methods designed to be used for accelerometers worn on the hip on individuals outside of controlled laboratory settings.

The 2011-2014 NHANES data collection uses a newer version of an accelerometer (ActiGraph GT3X+) and collected high frequency (80Hz) acceleration from a wrist-worn monitor. Advantages of this new protocol include the possibility of objective examination of sleep measures (18), and significant improvement in 7-day wear compliance over a hip worn accelerometer (24). However, it is not clear how these data will be processed to examine activity and sedentary behavior since there is limited research to develop validated algorithms for the trixaxial, high frequency accelerometer data from this wrist-worn sensor. The limited research includes work by Hildebrand et al. (13) that developed regression equations to estimate energy expenditure from a wrist-worn accelerometer and work by Zhang et al. (25) that develops methods to classify four activity types from wrist-worn data. The current investigation develops algorithms to estimate four measures of physical activity: (a) MET-hours, (b) time (min) in light (METs<3), moderate (3<=METs<6) and vigorous (METs>=6) activity, (c) time in sedentary activities versus not, and (d) time in locomotion versus not. We estimate locomotion time because locomotion is the primary mode by which most people accumulate moderate to vigorous activity. For each metric, we develop and compare both a machine learning method and simpler linear regression or decision tree methods. The latter methods could be implemented by a non-statistical expert using the ActiLife software and a spreadsheet. We also compare the MET-hour and activity intensity estimates for the wrist worn algorithm to the Freedson cut-point method (9) and to the refined Crouter two-regression method (6) for the hip worn accelerometer. The discussion contains a preliminary evaluation of the new wrist methods on two additional individuals engaged in several hours of free-living activity. These subjects are different from the n=20 on whom the new wrist methods were developed.

Methods

Participants and Data Collection:

Twenty healthy individuals between the ages of 20 and 39 years participated in this study. Participant characteristics are shown in Table 1. Prior to commencing the study protocol, participants completed a health history and a physical activity status questionnaire. Participants read and signed an informed consent document approved by the Institutional Review Board at the University of Massachusetts Amherst.

Each participant completed a treadmill routine and one of three activity routines comprised of sedentary, lifestyle and sport activities (Table 3). The two routines were completed in two separate visits. During each visit, participants wore an ActiGraph GT3X+ (ActiGraph Corporation, Pensacola, FL) secured on the dominant wrist with a velcro strap and a hip worn Actigraph GT3X accelerometer secured on the right hip with an adjustable belt. These accelerometers record gravity as well as movement. Oxygen consumption was measured breath-by-breath using the Oxycon mobile indirect calorimetry system (Carefusion, Yorba Linda, CA). All participants completed the treadmill routine during visit 1. Each treadmill speed was performed for six minutes followed by a three to four minute recovery break. The other activity routine was conducted during visit 2. These activities were also performed for six minutes each. The activities were performed as similar to free-living conditions (e.g., gardening and raking were performed outside, basketball was performed in a gym on a court) as possible. Subjects were seated exclusively throughout the driving and office work activities.

Instrumentation:

The ActiGraph GT3X+ monitor was secured on the dominant wrist with a velcro strap. This device is a small (4.6 cm x 3.3 cm x 1.5 cm), lightweight (19 g), triaxial accelerometer. The sampling frequency of the GT3X+ ranges from 30 to 100 Hz; an 80 Hz sampling frequency was used in this study which is the same sampling frequency used in the 2011-2014 NHANES accelerometer study.

The Oxycon mobile indirect calorimetry system (Carefusion, Yorba Linda, CA) was used to collect breath-by-breath respiratory gas exchange data. To calculate actual METs for each activity, the average oxygen consumption in ($ml \cdot kg^{-1} \cdot min^{-1}$) for minutes 3-5 was divided by 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$.

Statistical Methods

We developed the methods in two steps. First we computed statistical summaries of the accelerometer signals, and then we used those summaries as covariates in models to estimate aspects of physical activity and sedentary behavior. We describe each step in more detail below.

The summary statistics we use are listed in Table 2. The majority of these statistics have been used in previous work (7),(25),(19) where the motivations for these statistics are discussed. In addition, since we hypothesize that arm position may be related to activity, we use an angle of acceleration of the wrist worn accelerometer To compute the angle, we determined the axis that recorded -1g when the arm was hanging straight down and 1g when the arm was raised vertically. The other axes were zero. We computed the angle as arcsin(axis used / vector magnitude) / (pi/2). We describe this procedure in detail because the definitions of x, y, and z-axes are manufacturer, and even model, specific.

We note that the GT3X+ does not have a gyroscope in it, so the angle is relative to the accelerometer's axes, not relative to a line that is perpendicular to the ground. All statistics for the wrist models were computed from non-overlapping 15-second intervals of acceleration.

These summary statistics are then used as covariates in two groups of four models that each estimate aspects of physical activity. The models in the first group (linear regression and decision trees) are relatively simple as they will be easier for others to use, have fewer parameters, and the parameters are easier to interpret. The ones in the second group are more complex and potentially powerful machine learning models.

For the relatively simpler modeling approaches, we use multiple linear regression to estimate METs, and we use decision tree models (1) to classify activity intensity and the other two outcomes (sedentary or not and locomotion or not). The potential inputs to these models are the statistical summaries of acceleration that were discussed previously. We built the models by considering all possible models that used two inputs each, and we selected the models that achieved the lowest leave-one-subject out cross-validated estimates of mean squared error (for MET estimation) and misclassification error (for the other outcomes). We note that a separate left-out subject was used to estimate the reported model performance as is recommended in (11) section 7.10.2.

For the machine learning approach, we implemented three machine-learning models (neural networks, support vector machines, and random forests). These approaches have been used to estimate physical activity from accelerometers previously (7),(25),(19). These models are non-linear regression models that can flexibly represent a wide variety of relationships between covariates and outcomes. A technical description of these methods is outside the scope of this article, and more detail can be found in (14).

For the comparison to existing hip linear regression models, we used the ActiLife software to calculate the counts per second for the vertical axis. Subsequently, we used the published methods in (9) and (6) to estimate METs.

All statistical analyses, including the pre-processing to compute the statistical summary inputs and specific implementation of the statistical learning methods, were performed using R-software (www.r-project.org) (22). This software is available from the first author.

Results

Table 3 lists the activities, the associated mean (sd) METs, and the mean (sd) of each of the accelerometer summary statistics. The table also lists which activities are considered as sedentary or not, and which are considered as locomotion or not.

Cross-validated estimates of performance are in Figures 1a,b. The models use accelerometer statistics computed from non-overlapping 15 second windows. Windows with length between 5 and 60 seconds were considered, and results were similar for windows that were at least 15 seconds long.

The left panel in Figure 1a shows that the random forest, neural network, and wrist linear regression estimates of METs are not significantly biased. The support vector machine method result in statistically significantly ($p<0.05$) biased estimates, but the biases are not large. The hip method estimates are significantly biased. We note that lack of significant bias only indicates that the mean of the estimates is not significantly different from the mean of the measured values, and many estimation methods (including a single constant mean) are likely to be unbiased. The right panel of Figure 1a contains estimates of root mean squared error (rMSE) for each method. The random forest results in the smallest rMSE, followed by the support vector machine, the neural network, and the linear regression for the wrist and the hip methods. The wrist linear regression rMSE is about 30% larger than the random forest. Since rMSE is the square root of the estimate variability plus the squared bias of the estimate, and both methods are approximately unbiased, the larger rMSE for the linear regression method is due to greater variability in estimates.

Figures 2a and b show the estimated residuals for the linear regression and random forest estimates of METs. These figures indicate that all methods tend to overestimate METs when the actual METs are low and underestimate when actual METs are high, even for methods that are unbiased overall. These over- and underestimations suggest that there are other factors influencing the relationship between activity and METs that are not accounted for by wrist acceleration. This may also be an instance of regression to the mean.

Figure 3 shows the average MET estimate for random forests and linear regression (wrist) by activity and indicates a good agreement between estimates and the criterion measure on average. The smaller variability of the random forest can be seen as well.

Table 4 and the first panel of Figure 1b contain more detail about the performance of the classification estimates of MET level. The decision tree method estimates MET level and classifies sedentary behavior as well as the more sophisticated machine learning methods. The machine learning methods identify locomotion slightly better than the decision tree, but the difference is small. All methods for wrist data can identify MET level, sedentary time, and locomotion relatively well.

The machine learning methods are cumbersome to illustrate; instead they are available as R code and objects from the first author. The linear regression model to estimate METs from wrist acceleration is

$$\mathbf{METs = 1.89378 + 5.50821(sdvm) - 0.02705(mangle)}$$

where sdvm is the standard deviation of the vector magnitude for the interval and mangle is mean angle of acceleration relative to vertical on the device for the interval (see Table 2 for equation to compute mangle). We recommend that if the model estimates a MET level less than 1.0, a value of 1.0 should be used. In our dataset, this situation only occurred 0.7% of the time (27 out of 3660 intervals).

We present the decision tree classification models next, and the variable definitions are in Table 2. We use a tree model to estimate activity intensity (MET level) in one of three categories: light (METs<3), moderate (>=3, <6), and vigorous (>=6). The estimated model is below.

<u>Activity Intensity Estimation Algorithm:</u>

If sdvm ≤ 0.26 and mangle > -52, then Light.

If sdvm ≤ 0.26 and mangle ≤ -52, then Moderate.

If 0.26 < sdvm ≤ 0.79 and mangle > -53, then Moderate.

If 0.26 < sdvm ≤ 0.79 and mangle ≤ -53, then Vigorous.

If sdvm > .79, then Vigorous (for any mangle).

Below are the two decision-tree models to classify whether an interval is sedentary or not and whether it is locomotion or not.

Sedentary or Not Estimation Algorithm:

If sdvm $\leq$ 0.098 and p625 $\leq$ 0.138 then Sedentary.

If sdvm $\leq$ 0.062 and p625 $>$ 0.138 then Sedentary.

If 0.062 $<$ sdvm $\leq$ 0.098 and p625 $>$ 0.138 then Non-sedentary.

If 0.098 $<$ sdvm $\leq$ 0.148 and p625 $\leq$ 0.118 then Sedentary.

If 0.098 $<$ sdvm $\leq$ 0.148 and p625 $>$ 0.118 then Non-sedentary.

If sdvm $>$ 0.148 then Non-sedentary (for any p625).

Locomotion or Not Estimation Algorithm:

If fpdf $\leq$ 0.039 and mangle $>$ -53 then Not locomotion.

If fpdf $\leq$ 0.020 and mange $\leq$ -53 then Not locomotion.

If 0.020 $<$ fpdf $\leq$ 0.039 and mangle $\leq$ -53 then Locomotion.

If fpdf $>$ 0.039 and mangle $\leq$ -62 then Locomotion.

If 0.039 $<$ fpdf $\leq$ 0.060 and mangle $>$ -62 then Not locomotion.

If fpdf $>$ 0.060 and mangle $>$ -62 then Locomotion.

Figures 4a,b,c describe these models graphically, and can be interpreted as follows. One figure is displayed for each of the three classification models, and each figure has one of the statistics that is used to classify on the X-axis and another on the Y-axis. The accelerometer statistics from each time interval define a point in the panel, and each time interval is classified according to where it falls in the panel. The classification tree defines the boundary for each class. For instance, the MET level classification model uses the standard deviation of the vector magnitude (sdvm, x-axis) and the mean angle (mangle, y-axis) to classify each time interval. If a 15-second interval had a sdvm=0.1 gs and an angle of -30 degrees, it would be classified as Light. If the standard deviation remained at 0.1 and the angle decreased to -70 degrees, the boundary at -53 degrees would be crossed, and the interval would be classified as Moderate.

Discussion:

We have developed and evaluated statistical models to estimate aspects of physical activity from an ActiGraph GT3X+ that is worn on the wrist and collects triaxial data at 80 Hz. The models estimate MET-hours, time in different activity intensity categories (light, moderate, and vigorous), the amount of time the wearer is sedentary or not, and the amount of time the wearer is undertaking locomotion or not. We consider two types of statistical models: sophisticated machine learning models (neural networks, support vector machines, and random forests) and simpler methods (multiple linear regression and decision trees) that could be implemented in a spreadsheet. All models are available from the authors. As inputs, both sets of models use summaries of the accelerometer signals that can be obtained from ActiGraph ActiLife software.

This investigation provides further evidence that acceleration measurements from the wrist can be used to estimate energy expenditure accurately and relatively precisely. Starting with high frequency, triaxial wrist GT3X+ data, the sophisticated machine learning methods estimated energy expenditure more precisely than the multiple linear regression approach, but both the sophisticated methods performed quite similarly to decision tree methods for estimated intensity level. Cross-validated estimates of both sets of approaches to the wrist data were more accurate and precise than methods from Freedson et al (9) and Crouter et al (6) which use low frequency one-axis hip acceleration measurements. Those methods were developed on different datasets, and our methods were evaluated by cross-validation. That may explain some of the differences in performance. Additionally, we note that the primary purpose of the current study is to develop methods to estimate aspects of physical activity and inactivity from high-frequency triaxial wrist acceleration measurements. It is beyond the scope of this article to consider all of the pros and cons of measuring acceleration at various locations, sampling frequencies, and numbers of axes.

The new methods tend to overestimatie METs at lower MET levels and underestimate METs at the higher MET levels (Figures 2a and b). This is not surprising as wrist motion and acceleration is probably a relatively larger fraction of total activity at lower intensity levels and a relatively smaller one at higher intensity levels. As others have shown, accelerometer signals from sensors worn at multiple locations reduce these systematic errors (2). Additionally, both sets of analytics applied to the wrist data were able to detect both sedentary time and locomotion time. Algorithms to accurately detect locomotion and sedentary time from a wrist worn accelerometer may be useful and meaningful metrics for the NHANES accelerometer study to examine the relationship between these objectively measured behaviors relative to health outcomes. Our results suggest that when the right summaries of acceleration are used as inputs, useful estimates of PA and sedentary behavior can be made without reliance on “black-box” machine learning algorithms. One limitation of the current study is that acceleration was measured on the dominant wrist, rather than the non-dominant wrist that is used in NHANES. The effect of this difference on the validity of our estimates is unclear. Additionally, differences in accelerometer orientation may cause some acceleration summary statistics to be different depending on whether the accelerometer is worn on the right or left wrist. We believe that these issues warrant investigation. We note that (25) found little difference in classification accuracy using data from the left or the right wrist.

As second limitation is that our linear regression and decision tree models both only use two statistical summaries as inputs to calculate their estimates. While we chose to use two inputs in order to make the models visually interpretable (e.g. Figures 4abc), it is certainly possible that models with more (or even fewer) inputs would perform better. There is limited room for improvement in some of the classification tasks, but the there is more room for improvement in the estimate of METs and the classification of activity intensity.

An anonymous referee raised the question of whether our non-sedentary tasks included intervals of standing still and pointed out that classification of sedentary behavior (versus not) probably would be much easier if standing still was not included. While we do not have a video record to see exactly what each subject was doing, we did examine detailed plots of acceleration vector magnitude from the wrist over time for each subject and activity. These plots revealed that, similar to driving and office work, the golf, basketball, laundry, gardening, raking, vacuuming, and dusting activities all include stretches of time when the vector magnitude remains at 1g with only small variation. This suggests standing still, but in the future it is recommended that video recordings of free-living behavior be obtained to be able to identify standing time within each activity segment.

A final limitation is that the current investigation uses calibration data that were collected from a relatively small group of young subjects who completed a scripted set of activities. While these data cover a range of activities (including driving) and energy expenditure levels, several recent investigators have found that models estimated from lab-based data can perform poorly when applied to data from free-living people (8),(10). We preliminarily investigated this by applying the new methods to wrist data from two additional individuals who were directly observed by trained observers for two hours each. These new participants were free-living in the sense that they were not told what to do, and the observers recorded what these participants did using a protocol similar to one described in (16). Figure 5 summarizes the results. These methods and results are promising, but the question of how methods described in this investigation will generalize to free-living people requires further investigation.

Acknowledgments:

This project was supported by National Cancer Institute Grant R01 CA121005 at the University of Massachusetts, Amherst. Shai He’s work was supported by a gift by Joan Barksdale to support research experience for undergraduates in the Department of Mathematics and Statistics at the University of Massachusetts, Amherst. The authors thank three anonymous referees for comments that substantially improved this paper.

References:

1. **Albinali F, Intille S, Haskell W, and Rosenberger M**. Using wearable activity type detection to improve physical activity energy expenditure estimation. ACM, 2010, p. 311-320.
2. **Bao LaIS**. Activity recognition for user-annotated acceleration data. In: *Pervasive*, edited by F. FAaM2004, p. 1-17.
3. **Brage S, Wedderkopp N, Franks PW, Andersen LB, and Froberg K**. Reexamination of validity and reliability of the CSA monitor in walking and running. *Med Sci Sports Exerc* 35: 1447-1454, 2003.
4. **Crouter SE, Clowers KG, and Bassett DR, Jr.** A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol* 100: 1324-1331, 2006.
5. **Crouter SE, Flynn JI, and Bassett DR**. Estimating Physical Activity in Youth Using a Wrist Accelerometer. *Medicine and science in sports and exercise* 47: 944-951, 2015.
6. **Crouter SE, Kuffel E, Haas JD, Frongillo EA, and Bassett DR**. Refined Two-Regression Model for the ActiGraph Accelerometer. *Medicine and science in sports and exercise* 42: 1029-1037, 2010.
7. **Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, and Marshall S**. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological measurement* 35: 2191-2203, 2014.
8. **Freedson PS, Lyden K, Kozey-Keadle S, and Staudenmayer J**. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample. *J Appl Physiol* 2011.
9. **Freedson PS, Melanson E, and Sirard J**. Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.
10. **Gyllensten IC, and Bonomi AG**. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. *IEEE Trans Biomed Eng* 58: 2656-2663, 2011.
11. **Hastie T, Tibshirani R, and Friedman JH**. *The elements of statistical learning : data mining, inference, and prediction*. New York, NY: Springer, 2009, p. xxii, 745 p.
12. **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P**. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc* 32: S442-449, 2000.
13. **Hildebrand M, Van Hees VT, Hansen BH, and Ekelund U**. Age Group Comparability of Raw Accelerometer Output from Wrist- and Hip-Worn Monitors. *Medicine and science in sports and exercise* 46: 1816-1824, 2014.
14. **James G, Witten D, Hastie T, and Tibshirani R**. *An introduction to statistical learning : with applications in R*. New York: Springer, 2013, p. xvi, 426 pages.
15. **Lyden K, Keadle SK, Staudenmayer J, and Freedson PS**. A method to estimate free-living active and sedentary behavior from an accelerometer. *Medicine and science in sports and exercise* 46: 386-397, 2014.
16. **Lyden K, Petruski N, Staudenmayer J, and Freedson P**. Direct observation is a valid criterion for estimating physical activity and sedentary behavior. *Journal of physical activity & health* 11: 860-863, 2014.
17. **Mannini A, Intille SS, Rosenberger M, Sabatini AM, and Haskell W**. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise* 45: 2193-2203, 2013.

18. **Martin JL, and Hakim AD**. Wrist actigraphy. *Chest* 139: 1514-1527, 2011.

19. **Rothney MP, Neumann M, Beziat A, and Chen KY**. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol* 103: 1419-1427, 2007.

20. **Staudenmayer J, Pober D, Crouter S, Bassett D, and Freedson P**. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol* 107: 1300-1307, 2009.

21. **Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, and Ainsworth BE**. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise* 32: S450-456, 2000.

22. **Team RC**. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2014.

23. **Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, and McDowell M**. Physical activity in the United States measured by accelerometer. *Medicine and science in sports and exercise* 40: 181-188, 2008.

24. **Troiano RP, McClain JJ, Brychta RJ, and Chen KY**. Evolution of accelerometer methods for physical activity research. *British journal of sports medicine* 48: 1019-1023, 2014.

25. **Zhang S, Rowlands AV, Murray P, and Hurst T**. Physical Activity Classification using the GENEA Wrist Worn Accelerometer. *Medicine & Science in Sports & Exercise* 2011.

Figure Legends

Figure 1a: Relative performance of wrist methods and hip method to estimate METs. “Hip LR” is Freedson et al’s linear regression (10) and RC Hip 2-R is Crouter et al’s refined two-regression method (6).

Figure 1b: Relative performance of the wrist methods to classify MET level, sedentary time, and locomotion time.

Figure 2a: Residual plots of wrist and hip linear regression estimates of METs. Each point represents 15” for the wrist linear regression, 10” for the refined Crouter 2-regression method, and 60” for Freedson’s method.

Figure 2b: Residual plots of random forest, neural network, and support vector machine estimates of METs. Each point represents 15” of data.

Figure 3: Mean performance of wrist models to estimate average METs per activity.

Figure 4a: The tree model to classify activity intensity. For figures 4abc, each axis lists a summary statistic for the accelerometer signals. Those statistics are computed for each 15-second window that defines a point in on the plot. The region’s label is the estimated class.

Figure 4b: The tree model to classify sedentary time.

Figure 4c: The tree model to classify locomotion time.

Figure 5: Evaluation of the methods versus direct observation on two free-living subjects who are different from the subjects used to develop the models.

Tables:

Table 1: Participant characteristics: mean (sd).

n	20
Female / Male	10/10
Age, years	24.1 (4.5)
BMI, kg / m^2	23.9 (2.9)
Height, cm	170.4 (10.9)
Weight, kg	69.9 (14.7)

Table 2: Statistics used to summarize the accelerometer signals.

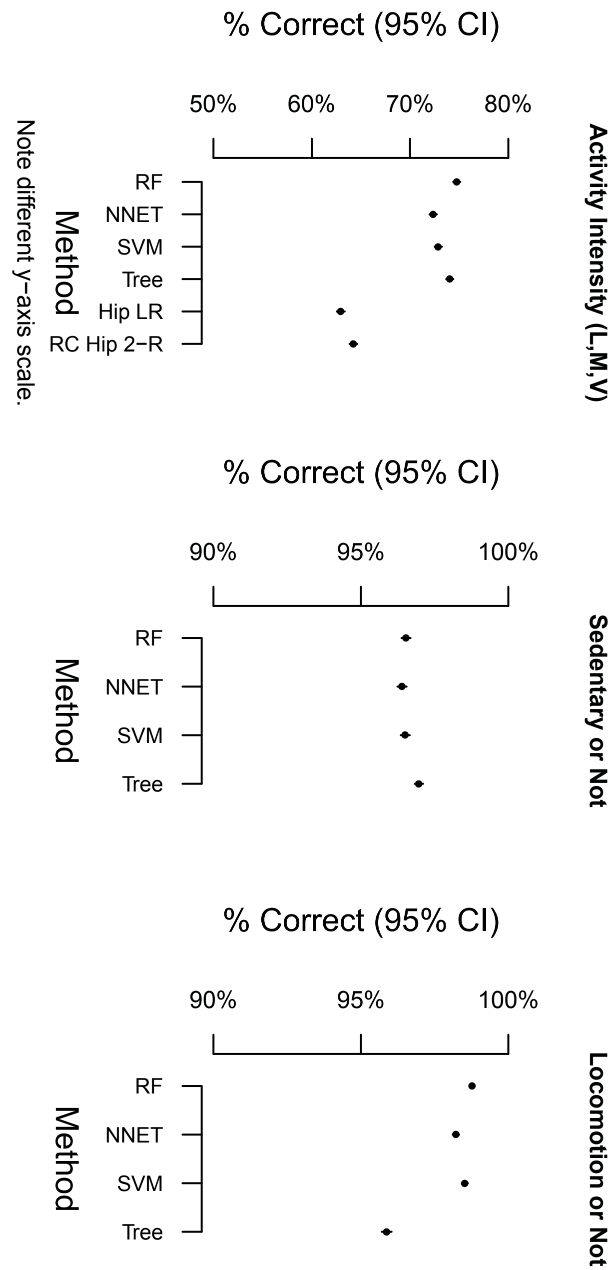
Statistic	**Definition**
Mean of vector magnitude (vm) (mvm)	Compute vm = sqrt(x^2+y^2+z^2). This statistic is the sample mean of vm in interval.
Standard deviation of vector magnitude (sdvm)	This statistic is the standard deviation of vm in interval.
Percentage of the power of the vector magnitude that is in 0.6 – 2.5 Hz (p625)	Estimate a spectral density of vm using a fast Fourier transform. Compute the modulus corresponding to each frequency. This statistic is the sum of those moduli divided by the sum of the moduli at each frequency.
Dominant frequency of vector magnitude. (df)	This statistic is the frequency that corresponds to the largest modulus.
Fraction of power in vector magnitude at dominant frequency. (fpdf)	This statistic is the modulus of the dominant frequency divided by the sum of the moduli at each frequency.
Mean angle of acceleration relative to vertical on the device. (mangle)	Compute the angle = 90arcsin(x/vm)/(pi/2). This statistic is the sample mean of those angles in the interval.
Standard deviation of the angle of acceleration relative to vertical on the device. (sdangle)	This statistic is the sample standard deviation of the angles in the interval.

Table 3: Summary statistics for each activity.

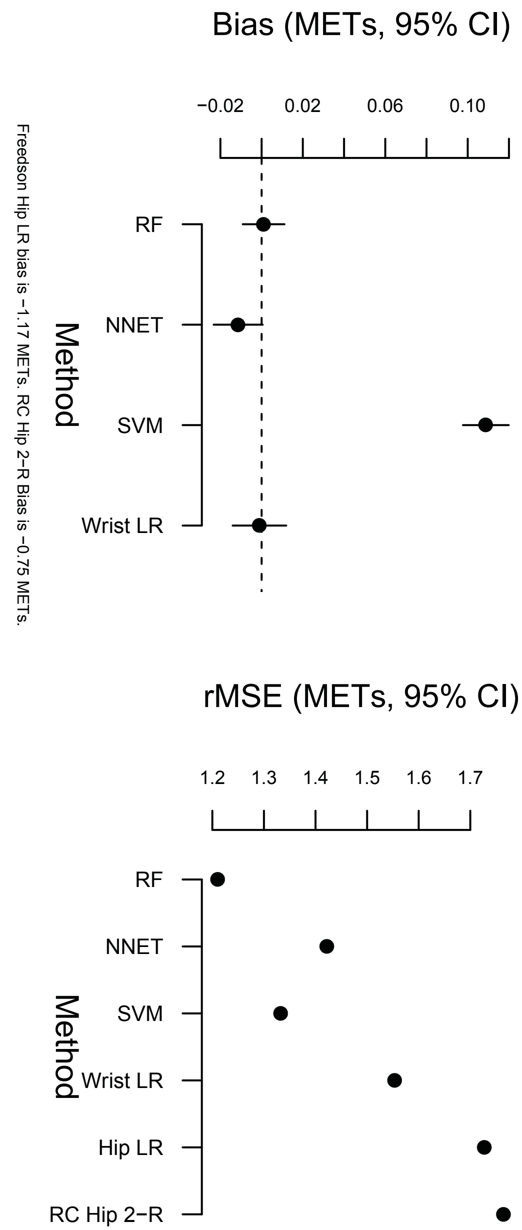
Activity	Locomotion?	Sedentary?	METs	Mean Vector Magnitude	Std Dev Vector Magnitude	Mean Angle	Std Dev Angle	% Power 0.6-2.5Hz	Dominant Freq	Power at Dom Freq / Total Power
Treadmill 5.5mph	Yes	No	9.21 (0.78)	1.82 (0.23)	1.21 (0.22)	-5.23 (5.23)	48.22 (4.07)	0.16 (0.08)	2.74 (0.55)	0.14 (0.03)
stairs	Yes	No	8.08 (0.43)	1.10 (0.06)	0.36 (0.10)	-62.46 (10.07)	13.64 (7.28)	0.36 (0.09)	2.00 (0.57)	0.10 (0.04)
Treadmill 4.0mph 5%	Yes	No	7.54 (0.81)	1.25 (0.17)	0.37 (0.19)	-62.27 (19.99)	13.63 (11.92)	0.36 (0.09)	2.12 (0.57)	0.15 (0.05)
Tennis	No	No	6.57 (1.92)	1.24 (0.17)	0.58 (0.30)	-41.65 (12.74)	24.21 (6.58)	0.30 (0.08)	1.17 (0.78)	0.06 (0.02)
Shooting Basketball	No	No	6.57 (1.89)	1.38 (0.16)	0.78 (0.24)	-30.35 (17.65)	32.97 (7.37)	0.26 (0.05)	1.33 (0.68)	0.06 (0.01)
Treadmill 3.0mph 5%	Yes	No	5.28 (0.49)	1.07 (0.05)	0.22 (0.04)	-67.17 (11.83)	7.50 (5.57)	0.42 (0.06)	1.80 (0.26)	0.15 (0.04)
Fast walk	Yes	No	5.20 (1.10)	1.19 (0.11)	0.30 (0.06)	-70.54 (3.78)	10.56 (2.91)	0.42 (0.05)	1.78 (0.45)	0.14 (0.05)
Stacking Boxes	No	No	4.99 (0.57)	1.04 (0.02)	0.25 (0.07)	-6.99 (20.76)	22.57 (10.26)	0.28 (0.06)	1.63 (0.51)	0.07 (0.03)
Raking	No	No	4.25 (0.84)	1.09 (0.07)	0.32 (0.15)	-10.94 (28.95)	20.91 (7.01)	0.22 (0.04)	1.15 (0.61)	0.05 (0.01)
Walking Carrying groceries	Yes	No	4.19 (0.58)	1.03 (0.02)	0.27 (0.05)	-66.43 (8.09)	10.38 (1.35)	0.43 (0.06)	1.90 (0.10)	0.16 (0.04)
Self-paced walk	Yes	No	3.84 (0.60)	1.10 (0.07)	0.23 (0.04)	-70.60 (4.49)	7.60 (3.53)	0.43 (0.06)	1.62 (0.33)	0.15 (0.05)
Golf (swinging)	No	No	3.57 (0.70)	1.13 (0.20)	0.40 (0.41)	-47.17 (16.38)	19.57 (12.22)	0.35 (0.10)	1.34 (0.46)	0.08 (0.04)
Slow walk	Yes	No	3.31 (0.44)	1.06 (0.05)	0.18 (0.03)	-70.50 (3.11)	6.02 (2.87)	0.39 (0.07)	1.52 (0.29)	0.14 (0.05)
Vacuuming	No	No	3.13 (0.40)	1.03 (0.03)	0.13 (0.07)	-31.40 (34.56)	12.75 (5.72)	0.19 (0.06)	1.47 (2.23)	0.04 (0.01)
Gardening	No	No	2.87 (0.41)	1.06 (0.07)	0.29 (0.19)	-28.51 (31.33)	17.26 (5.36)	0.18 (0.07)	2.53 (1.47)	0.04 (0.01)
Laundry	No	No	2.42 (0.56)	1.04 (0.03)	0.22 (0.09)	-1.33 (22.03)	24.56 (9.86)	0.27 (0.07)	1.59 (1.13)	0.05 (0.01)
Dusting	No	No	2.42 (0.36)	1.03 (0.03)	0.12 (0.04)	-17.75 (24.40)	14.03 (6.45)	0.24 (0.08)	2.03 (0.88)	0.05 (0.02)
Driving	No	Yes	1.47 (0.16)	1.03 (0.03)	0.08 (0.06)	21.85 (16.73)	5.31 (4.56)	0.13 (0.06)	7.32 (7.67)	0.03 (0.02)
Office work	No	Yes	1.20 (0.09)	1.01 (0.02)	0.02 (0.02)	11.42 (10.21)	1.90 (1.48)	0.07 (0.04)	5.84 (3.36)	0.02 (0.01)

Table 4: MET level classification confusion matrices.

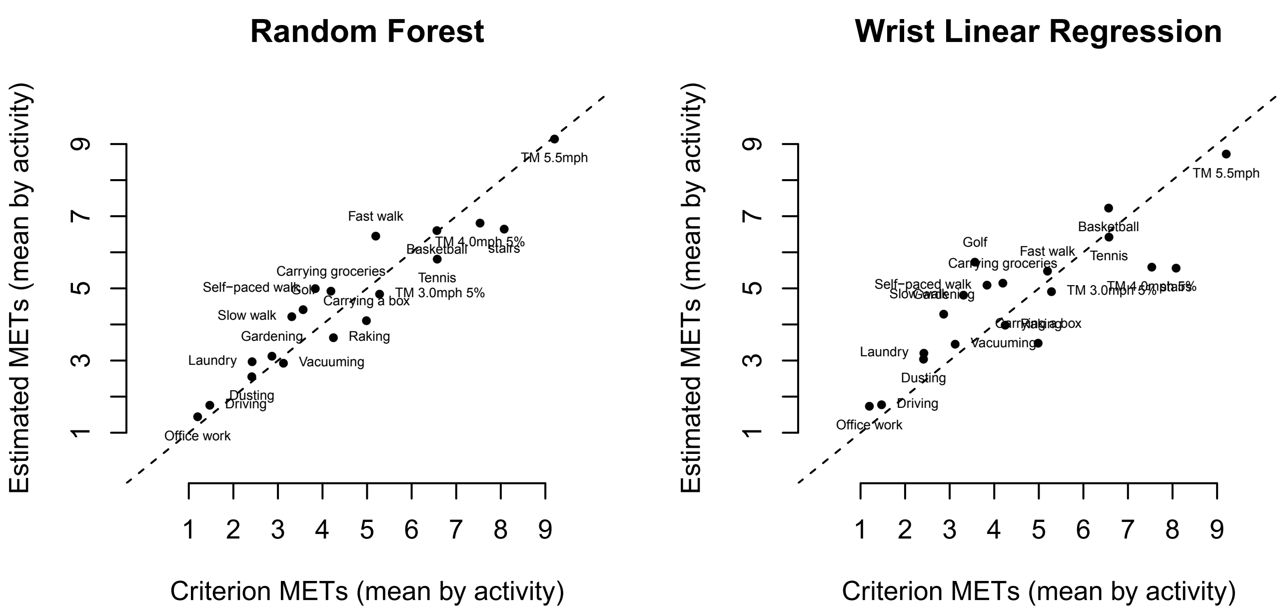
Random Forest		Estimate		
		light	moderate	vigorous
True Category	light	698	230	2
	moderate	239	1088	202
	vigorous	1	237	963
Tree		Estimate		
		light	moderate	vigorous
True Category	light	718	193	19
	moderate	229	987	313
	vigorous	2	194	1005

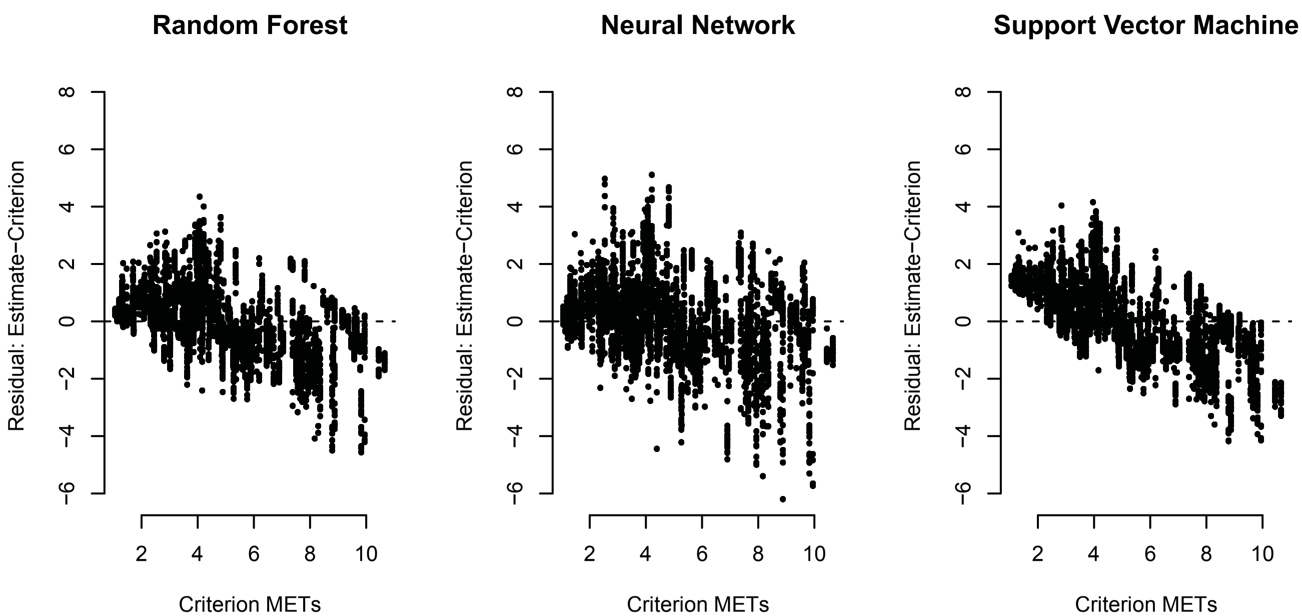


% Correct (95% CI)
50%
60%
70%
80%
RF
NNET
SVM
Tree
Hip LR
RC Hip 2−R
Method
Activity Intensity (L,M,V)
Note different y−axis scale.
% Correct (95% CI)
90%
95%
100%
RF
NNET
SVM
Tree
Method
Sedentary or Not
% Correct (95% CI)
90%
95%
100%
RF
NNET
SVM
Tree
Method
Locomotion or Not

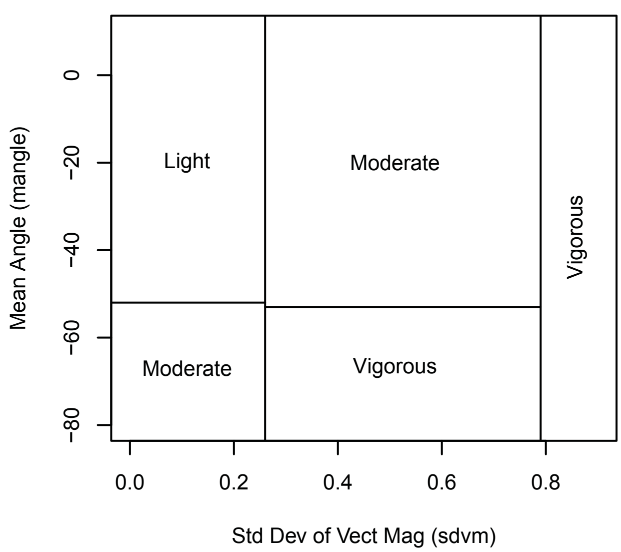

Bias (METs, 95% CI)
−0.02
0.02
0.06
0.10
RF
NNET
SVM
Wrist LR
Method
Freedson Hip LR bias is −1.17 METs. RC Hip 2−R Bias is −0.75 METs.
rMSE (METs, 95% CI)
1.2
1.3
1.4
1.5
1.6
1.7
RF
NNET
SVM
Wrist LR
Hip LR
RC Hip 2−R
Method

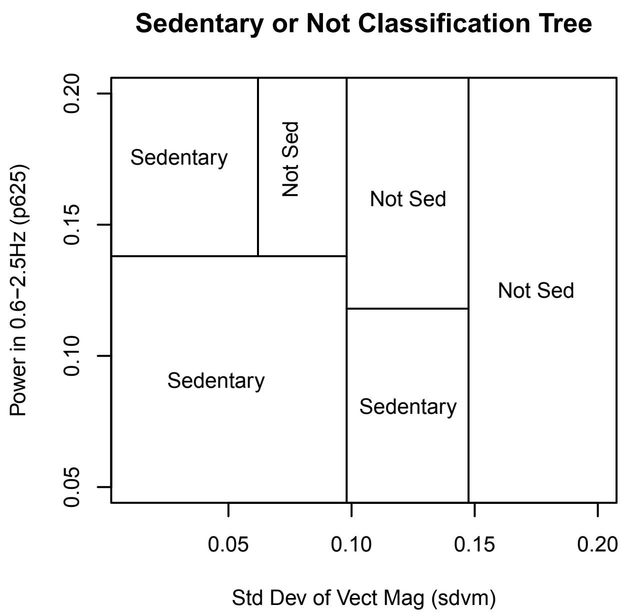

Linear Regression Wrist
Freedson Regression Hip
Refined Crouter 2−Reg Hip
Residual: Estimate−Criterion
5
0
−5
2
4
6
8
10
Criterion METs

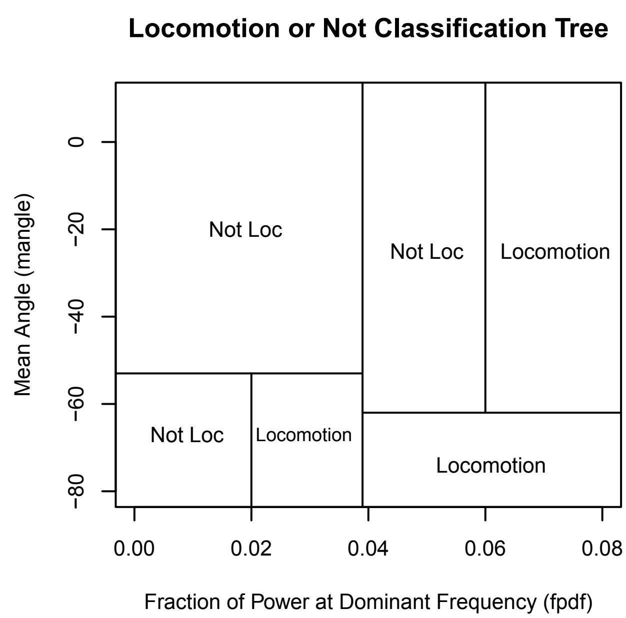

Random Forest
Wrist Linear Regression
Estimated METs (mean by activity)
Criterion METs (mean by activity)
TM 5.5mph
Fast walk
TM 4.0mph 5%
Basketball
stairs
Tennis
Carrying groceries
Self−paced walk
Golf
TM 3.0mph 5%
Carrying a box
Slow walk
Gardening
Raking
Laundry
Vacuuming
Dusting
Driving
Office work


Random Forest
Neural Network
Support Vector Machine
Residual: Estimate−Criterion
Criterion METs
8
6
4
2
0
−2
−4
−6
2
4
6
8
10

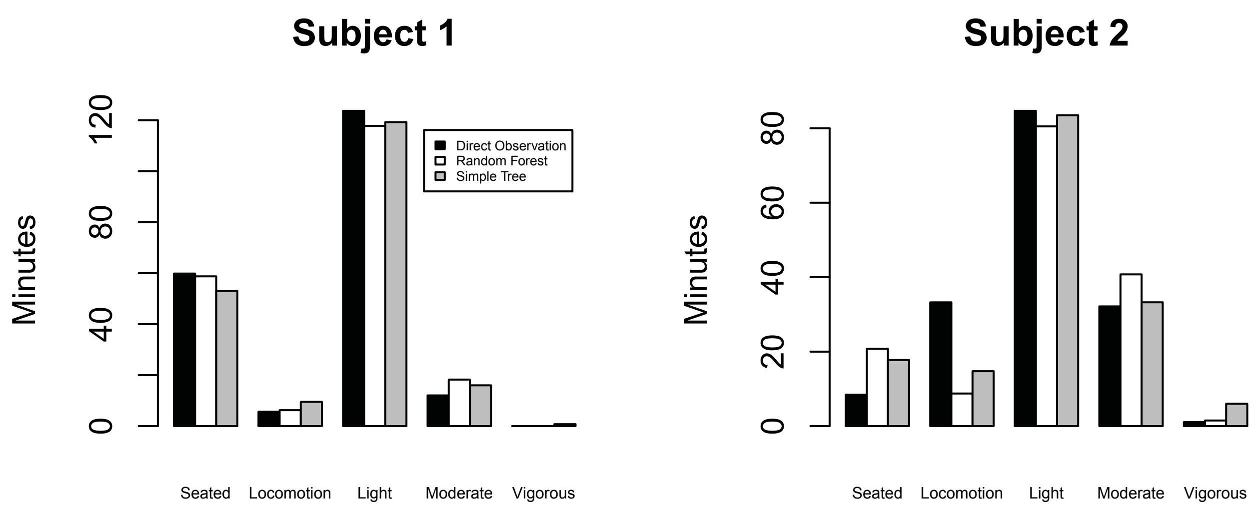
Activity Intensity Classification Tree




Sedentary or Not Classification Tree
Sedentary
Not Sed
Not Sed
Not Sed
Sedentary
Sedentary
Power in 0.6−2.5Hz (p625)
0.05
0.10
0.15
0.20
Std Dev of Vect Mag (sdvm)



Locomotion or Not Classification Tree
Not Loc
Not Loc
Locomotion
Not Loc
Locomotion
Locomotion
Mean Angle (mangle)
0
−20
−40
−60
−80
0.00
0.02
0.04
0.06
0.08
Fraction of Power at Dominant Frequency (fpdf)



Subject 1
Subject 2
Minutes
0
40
80
120
0
20
40
60
80
Direct Observation
Random Forest
Simple Tree
Seated
Locomotion
Light
Moderate
Vigorous