

## MODELING OF ORDINAL LONGITUDINAL ACCELEROMETER DATA

DREW M. LAZAR\*

*Department of Mathematical Sciences, Ball State University, Muncie, IN 47304*  
*Email: dmlazar@bsu.edu*

FATIMAH ALAWAD

*Department of Mathematical Sciences, Ball State University, Muncie, IN 47304*  
*Email: fkalawad@bsu.edu*

MUNNI BEGUM

*Department of Mathematical Sciences, Ball State University, Muncie, IN 47304*  
*Email: mbegum@bsu.edu*

ALEXANDER H.K. MONTOYE

*Integrative Physiology and Health Science, Alma College, Alma, MI 48801*  
*Email: montoyeah@alma.edu*

### SUMMARY

Improvements in accelerometer technology has led to new types of data on which more powerful predictive models can be built to assess physical activity. This paper implements an ordinal random forest model with recursive forecasting to take into account the ordinal, longitudinal nature of responses. The data comes from 28 adults performing activities of daily living in two visits, while wearing accelerometers on the ankle, hip, right and left wrist. The first visit provided training data and the second testing data so that an independent sample, cross-validation approach could be used. For this type of data, prior responses are not available at the testing stage or in practice. However, recursive forecasts can be made with prior predictions in place of lagging responses on models which were built to use lagging responses as explanatory variables. Models are fit to account for multiple time series, with different time series for each participant in the study. We found that ordinal random forest, when the time series is taken into account, produces better accuracy rates and better linearly weighted kappa values than both ordinary ordinal forest and random forest. On the testing set, the lowest error rates were produced by the ankle (28.0%), followed by the left wrist (28.7%), hip (28.9%) and then the right wrist (30.2%) using the best performing decision model for a four-activity level response. In addition, linearly weighted kappa values indicated substantial agreement. The approach of this work can be adapted to other types of longitudinal ordinal models for improvements in modeling techniques.

*Keywords and phrases:* Longitudinal Data, Multiple Time Series, Accelerometers, Ordinal Regression, Ordinal Forest

*AMS Classification:* 62M10

---

\* Corresponding author  
© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

# 1 Introduction

Accelerometers are devices that track acceleration forces in three orthogonal directions. They have been used to objectively assess levels of physical activity since the first physical activity sensor based on accelerometers were developed in the 1980s. Since then, the power, battery space, size, and accuracy of the hardware of these accelerometer-based physical activity sensors (which we now refer to as accelerometers) has improved dramatically. Recently, researchers have been able to view and use raw signal acceleration data instead of relying on count thresholds to predict physical activity responses (Troiano et al., 2014). This raw data is large, as it contains acceleration signals at 30-100 measurements per second along the x, y, and z axes. Therefore, there are computational difficulties in processing physical activity information reliably and accurately from raw signal data for acceleration.

One method that has been implemented is to summarize raw acceleration data into every-30-second epochs (Montoye et al., 2017b,a; Lazar et al., 2020). These summaries make up new explanatory variables and consist of summaries for each axis of acceleration and summaries of pair-wise associations between these axes. The summaries consist of the mean, variance, minimum, limits, quartiles and the pair-wise correlation coefficients between the x and y, x and z and y and z axes. This new, summarized explanatory data, along with the level of measured or observed activity level in each 30 second epoch, can be used to create statistical decision models. Decision trees, random forests, boosting and bagging, parametric linear and nonlinear models, and neural networks, among others, are types of models used for this purpose (Montoye et al., 2017a, 2018; De Vries et al., 2011).

The activity level response in our data has four ordered classes, sedentary behavior (SED), light physical activity (LPA), moderate physical activity (MPA) and vigorous physical activity (VPA) and we use different methods to predict the response taking this into account. Specifically, the level of activity that a participant engages in during a particular period informs the level of activity that the participant engages in subsequent periods. For example, if a participant engages in LPA in a particular epoch, the participant is more likely to engage in LPA or MPA in the next epoch than VPA. Thus, we account for the temporal, time-series nature of the responses. As there are multiple participants' activity levels represented in our data set, we need to account for multiple time series in our models. In addition, taking into account ordinality is a crucial consideration in building statistical models (Brant, 1990). For our data, we consider a modification of random forest known as ordinal forest that takes into account the ordinality of the responses.

Accelerometer data is often multiple time series, sampled over periods of exercise and movement for different participants. At the same time, the responses in training sets are often measured in ordinal classes of activity levels. Our motivation for this research is to integrate these two important features in order to improve the accuracy and the quality of models of accelerometer data. This is a novel approach to modeling this type of data, and these methods can be improved and extended to data from other physiological studies and to data from other domains as well.

## 2 Background

Accelerometers have been used to objectively measure levels of physical activity since the 1980's. During the 1990's the use of accelerometers for this purpose increased greatly. From 1981 to 1996 there was an average of 10 articles per year on the use of accelerometers in physical activity research, and by 2013-2014 this average had increased to an average of 600 articles per year (Troiano et al., 2014).

Over this time period, there has been steady movement from using activity count thresholds, to using sums of acceleration signals over epochs, to using the raw acceleration signals in three axes for analysis and prediction of the level of activity engaged in by a wearer of an accelerometer (Acti-GraphCorp, 2017).

There has been extensive literature using this latter approach. These use a number of different types of statistical and prediction models as well as analyze different placements, different ways of measuring level of activity and specific uses of accelerometers in predicting physical activity levels. For example, ensemble methods based on decision trees such as bagging, boosting and random forest have been utilized in Montoye et al. (2018), Ellis et al. (2014), Lazar et al. (2020) and Pavey et al. (2017), among others. Parametric and linear models to analyze accelerometer data, including interpretation of coefficients and significance of variables, have been explored in Montoye et al. (2017a), Lazar et al. (2020) and Robert et al. (2009). Neural networks and deep learning have also been employed in De Vries et al. (2011), Nawaratne et al. (2020) and Montoye et al. (2017a). The longitudinal, time series nature of accelerometer data has been considered in modeling and prediction. Tan et al. (2019) employed long short-term memory (LSTM) networks to learn order dependence in a sequence predictions from accelerometer data. Hidden Markov models, which make Markov assumptions about the sequence of observable activity levels in practice and incorporate the influence of the latent activity levels in the unobservable states, are considered in Witowski et al. (2014), Nickel et al. (2011) and Leos-Barajas et al. (2017).

Amato et al. (2024) and references therein presents methods based on finite mixtures of matrix-variate distributions to address the clustering problem of longitudinal ordinal data. They considered a model based clustering algorithm with the assumption that the ordinal variable can be represented as a discretized version of an underlying latent continuous variable.

Recursive forecasting with prior predictions has been similarly implemented in De Vries et al. (2011) as "lag-one autocorrelation" in neural networks and Ellis et al. (2014) implemented it in random forest. In both cases, these produced improved prediction accuracy. The background and implementations of this approach is seen in Branch and Evans (2006), Ng and Young (1990) and Tyrallis and Papacharalampous (2017).

Considering the ordinality of responses and weights according to different types of misclassification is an idea that was introduced over thirty six years ago in Breiman et al. (1984) with the first formulation of decision trees. Recently, new ways of integrating the order of responses in ensemble methods have been introduced, such as in Hornung (2019) and Janitza et al. (2016). These methods show slight improvements in error rates for large data sets but improvements in assessment measures that account for ordinality. This is the case where the ordinal forest method in Hornung (2019) is applied in Hossain et al. (2021) to accelerometer data to predict a four-level ordered activity response.

Sasaki et al. (2016) notes overall error rates of classification algorithms of free-living physical activity from explanatory accelerometer data above an “acceptable” rate of 20%. Using recursive forecasting and taking into account previous prior predictions in testing and in practice in ordinal models, is meant to address such error rates and to consider and incorporate all relevant information into building prediction models.

### 3 Methodology

We used data collected from two separate visits to the Ball State Clinical Exercise Physiology Laboratory. These visits were designed to produce training and testing sets, respectively, for model building and testing. We analyzed this data accordingly, with random forest models, ordinal forest models and recursive forecasting models with one, two and three lag periods.

#### 3.1 Accelerometer Data

The data that we used consists of 29 variables, 28 of which are explanatory. The single response variable is the physical activity intensity level of the participant with possible values of SED, LPA, MPA and VPA. Four of the explanatory variables are time-invariant demographic information of the participants (age, height, weight, sex) and 24 features are summary statistics of raw signal accelerometer data summarized into every 30-second epochs. The data along each axis consists of summaries for each axis (mean, variance, minimum, maximum, and upper percentiles) and summaries of pair-wise correlations between these axes. Participants wore accelerometers on the hip, left and right wrists and on the ankle. By applying our models to a data set with the same variables that is collected independently from the training set, i.e., by using a test set, the predictive ability of these models is evaluated.

For the response variable, the 2011 Compendium of Physical Activities (Ainsworth et al., 2011) provided MET (metabolic Equivalent of Task) values for each activity. Intensities were set in four categories as:  $\leq 1.5$  METs as SED, 1.6-2.9 METs as LPA, 3.0-5.9 METs as MPA, and  $\geq 6$  as vigorous VPA. Observed activities and their intensities according to the MET scale served as the ground truth for development of prediction models using accelerometer data.

#### 3.2 Participants and Data

There were 30 participants in the study who provided the accelerometer data that is analyzed in this paper. Two separate visits were made by each of these participants to the Ball State Clinical Exercise Physiology Laboratory. These visits were planned to generate training and testing sets for model development and research. There were no orthopedic restrictions on any of the 30 participants. Ten adults of five females and five males were drawn from each of the three groups of ages 18-39, 40-59, and 60-79. In these age groups, there is a well-established variability in activity levels (Caspersen et al., 2000; Sallis, 2000) and this distribution of participants allows models to be developed for the general adult population. The distribution of the 1408 observations in the SED class in our training data was 27.8 percent in 18-39, 34.9 percent in 40-59, and 37.3 percent in 60-79, and the

distribution of the 298 observations in the VPA class was 50 percent in 18-39, 37.6 percent in 40-59 and 12.4 percent in 60-79. Overall, each of the the associations between age group and activity level, BMI (from height and weight) level and activity level, and sex and activity level were statistically significant in the training set according to a chi-squared test at significance level  $\alpha = 0.05$ . Two participants had invalid data during data collection, resulting in the data set for our sample of 28 individuals. Table 1 displays the participants' demographic statistics.

Table 1: Demographics of Subjects

	Total Sample	Male $n = 14$	Female $n = 14$
Age (years)	48 (19.6)	48.5 (19.83)	47.6 (20.15)
Weight (in lbs)	176 (34.72)	194.96 (27.53)	157.3 (31.32)
Height (in inches)	68.5 (3.52)	71 (2.62)	65.9 (2.3)
BMI	26.4 (4.16)	27.03 (3.07)	25.8 (5.07)

In Table 2, we see the distribution of activity levels in the next epoch given activity levels in the previous epoch for our 28 participants in the training set. For example, we see that when a participant is engaged in sedentary activity, then in 88.89% of the next epochs they are engaged in SED activity, 10.94% of the next epochs they are engaged in LPA, 00.27% of the next epochs they are engaged in MPA, and 00.00% of the time they are engaged in VPA. In Figure 1, we see a typical autocorrelation plot for a participant in the study, in this case, a 47 year old male, with a BMI (Body Mass Index) of 34.23. The horizontal access gives the lag in the series of activity level responses (coded 1 through 4) and the vertical access gives the correlation between observations with that lag. We see significant (outside the dashed line), but steadily decreasing correlation as the lag period is increased. In table 3, the lag correlations for the first 9 lags in the training set are computed for all participants in the training set, and the average correlations are taken.

Table 2: Distribution of the next activities (Row Percentages in Parenthesis)

Activity Level	SED	LPA	MPA	VPA
SED	1307 (88.89*)	161 (10.94)	4 (00.27)	0 (00.00)
LPA	139 (9.96)	1048 (75.13)	207 (14.84)	1 ( $\approx 0$ )
MPA	7 (00.63)	183 (16.38)	877 (78.51)	50 (04.48)
VPA	18 (05.56)	29 (08.95)	3 (00.93)	274 (84.57)

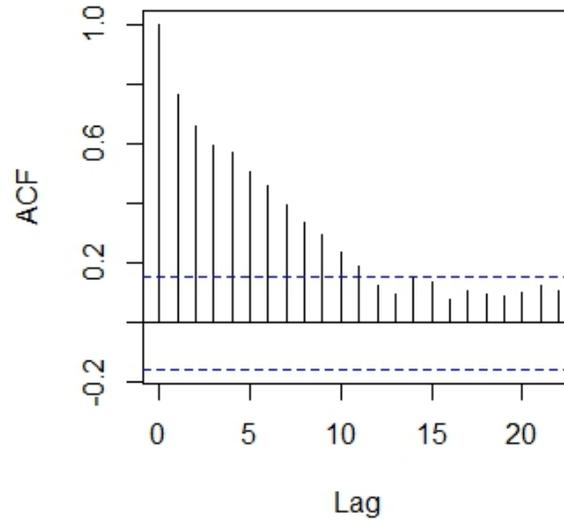


Figure 1: Autocorrelation Plot: Male, Age 47, BMI of 34.23

Table 3: Average Lag

Lag	0	1	2	3	4	5	6	7	8	9
AC	1.000	0.769	0.643	0.560	0.484	0.421	0.367	0.322	0.289	0.269

### 3.3 Protocol

Two exercise laboratory visits, visit 1 followed by visit 2, were used to collect accelerometer data. Visit 1 generated the training data for the construction of our models and visit 2 generated the test data for assessment.

In visit 1, participants were closely supervised by research staff and planned to provide a variety of activities of different intensities and speeds to fully train the model. Participants completed 11 tasks, beginning with 10 minutes of lying on a padded table. Then from Table 4, 10 tasks were allocated. Two from the sedentary category were selected for each participant, four from the category of lifestyle/chore activity, and four from the category of ambulatory/exercise were selected. These activities were selected at random and in such a way that about the same number of participants performed all the activities under each of the categories. For five minutes, and the task was performed, with the order of activities progressing from sedentary to lifestyle/chore to out-

Table 4: Activities during visit

<b>Sedentary</b>	<b>Lifestyle/Chore</b>	<b>Ambulatory/Exercise</b>
Reading, using a computer, watching television, writing, playing cards.	Standing, dusting, making a bed, folding laundry, sweeping, vacuuming, simulated gardening, picking up items from the floor.	Slow/fast overground walking, treadmill walking, overground jogging, treadmill jogging, stationary cycling, ascending stairs, descending stairs.

patient/exercise. Two from the sedentary category were selected for each participant, four from the category of lifestyle/chore activity, and four from the category of ambulatory/exercise were selected. These activities were selected at random and in such a way that about the same number of participants performed all the activities under each of the categories. For five minutes, and the task was performed, with the order of activities progressing from sedentary to lifestyle/chore to outpatient/exercise. Between operations, there was a 1-2 minute pause. Participants were asked to conduct sedentary exercises and lifestyle/chore operations as they would in their everyday lives. They were expected to maintain a consistent pace and intensity for the ambulatory/exercise activities.

In visit 2, study staff provided less structure than visit 1, with participants engaging in activities in the way they would in their everyday lives. This method has been previously used for generating testing data to assess the generalizability of models to free-living settings (Sasaki et al., 2016; Montoye et al., 2017b). Sixteen activities were performed by participants, each conducted for two to fifteen minutes. The participants were instructed to choose four activities from the sedentary group, four from the category of lifestyle/chore, and four from the category of outpatient/exercise. As previous research suggests that adults are often engaged in sedentary activity Montoye et al. (2017a); Statistics (2008), participants were asked to perform at least 40 minutes of sedentary activity. Unlike visit 1, visit 2 participants were able to select the time they spent performing activities, the activities to perform, and the order in which to perform them, within a structure.

In both visits, each participant wore four ActiGraph GT9X Connect accelerometers on the left and right wrists, right hips, and right ankles. The standard positioning was on the hip with count-based activity level estimation, but new placements have increasingly been more widely used, like the ankle and wrist. For better enforcement, particularly when worn on the wrist, and for enhanced ability to monitor such movement metrics, including measures, when worn on the ankle, these additional positions are selected. Models designed using thigh accelerometer data have demonstrated greater accuracy than hip or wrist-worn accelerometers, but less convenience and lower conformity with the location of the thigh accelerometer have also been recorded. At a rate of 60 samples per second, the accelerometers were initialized to collect acceleration data along the x, y, and z axes. Data were summed with characteristics such as mean, variance, minimum, maximum, and 70th, 80th,

and 90th percentile, and pair-wise axis associations over each epoch, in 30-second non-overlapping epochs. The demographic variables were also reported in the data for each participant's sex, age, height, and weight.

MET (Metabolic Equivalent of Task) values were observed by study staff for each activity according to the 2011 Compendium of Physical Activities Ainsworth et al. (2011). Intensities of operations were classified in four categories:  $\leq 1.5$  METs as sedentary (SED), 1.6-2.9 METs as mild (LPA), 3.0-5.9 METs as moderate (MPA), and  $\geq 6$  as vigorous (VPA). In both visits, participants took one to two minutes of rest at the end of each task before beginning the next activity. For the creation of prediction models using accelerometer data, these observed events and their intensities according to the MET scale served as the ground reality. This data was inserted into the data collection in accurate 30-second epochs until ground truth data were coded according to an operation speed.

## 4 Models

### 4.1 Random Forest

We implemented the random forest algorithm from the R `randomForest` package (Breiman et al., 2018). We used common specifications of the numerous features of random forest when constructing these models. Random forest requires that many decision trees are created from repeated bootstrap samples from the data. We took these bootstrap samples with replacement and when creating decision trees, we choose splits at nodes through the Gini impurity index (Tangirala, 2020). To avoid overfitting and undue influence of particular variables, a subset of explanatory variables is chosen at random for creating each tree. The number of such variables used for each tree was set as the square root of the number of variables. Majority voting was used to make decisions from new observations in testing. The splitting criteria, and decision voting of the random forest models, did not account for the ordinal nature of the responses.

### 4.2 Ordinal Forest

We implemented the ordinal forest algorithm from the R package `ordinalForest` (Hornung, 2019). Ordinal responses in our accelerometer data can be assigned numeric values and regression forests (Breiman, 2001) can be used for prediction. As quantitative data is at a higher level of measurement than ordinal data, this would account for ordinality, but because of the arbitrariness of these assignments, this approach is shown to not lead to improvements in model quality (Hornung, 2019). Instead, as summarized in Hossain et al. (2021), ordinal forest proceeds as follows:

1. Choose a large number of random, heterogeneous partitions of  $[0, 1]$  by  $J$  intervals, where  $J$  is the number of classes in the response ( $J = 4$ , in our case).
2. Represent each of the  $J$  classes by the midpoints of respective, ordered intervals in the partition.

3. For each representation of classes in the response variable by midpoints in 2., build a regression forest.
4. The representations used in the random forests built in 3. with the smallest out-of-bag errors are averaged. The regression forest built on that summary is the resulting model from ordinal forest.

In computing out-of-bag errors in step 3. above, different performance functions can be chosen, each of which emphasize different objectives of the model. We used the “proportional” performance function, which attempts to lower error rates from the larger classes at the expense of smaller ones and thus, attempts to lower the overall error rate.

### 4.3 Recursive Forecasting with Multiple Time Series

As shown in section 3.2, physical activity levels inform later physical activity levels for each participant. Thus, we use recursive forecasting to make predictions (Taieb et al., 2012).

Models created from accelerometer data are used to make assessments of physical activity levels without direction observation (Ward et al., 2005). Thus, we do not use a direct time series approach, because in testing, only accelerometer data and demographic information and not prior physical activity levels, are available for making predictions. Instead, prior predictions of activity levels, along with accelerometer information and demographic information, are used to forecast activity levels.

In addition, each participant engages in activity, independent of every other participant and has their own longitudinal measurements of activity levels. Thus, we build our recursive forecasting models for each participant, and in testing we take weighted averages of predictions of each model’s predictions, with the weights informed by variable importance and similarity in demographic variables of the testing and training observations.

#### 4.3.1 Ordinal Forest Models with Lagging Variables

For each participant  $h$  in the training set, we fit an ordinary, ordinal forest model and  $\ell$  time series models with ordinal forest. For time series model  $k = 1, \dots, \ell$ , we use  $k$  lagging values of the response variable as  $k$  additional explanatory variables. From training participant  $h$ , we denote the ordinal forest model that uses no lagging variables by  $of_{h,0}$  and the ordinal forest model that uses  $k$  lagging periods by

$$of_{h,k} \text{ for } k = 1, \dots, \ell.$$

#### 4.3.2 Recursive Prediction by Model Voting

For participant  $r$  in the testing set and for  $\ell$  lagging periods we make predictions by recursive forecasting. For  $k = 1, \dots, \ell$  we make weighted predictions from

$$of_{1,k}, \dots, of_{T,k}$$

where  $T$  is the number of training participants. We denote the models that produce weighted predictions for test participant  $r$  as  $w_r of_k$  for  $k = 1, \dots, \ell$ . The weights that  $w_r of_k$ , for  $k = 1, \dots, \ell$ , uses to make predictions from models  $of_{1,k}, \dots, of_{T,k}$  are determined as in section 4.3.3.

To make predictions for the  $n$  observations of participant  $r$  in the testing set we proceed as follows.

1. Add the “pseudo-lagging” variables  $sl_1, \dots, sl_\ell$  to the testing data, which will have as values prior predictions from our models  $w_r of_0, w_r of_1, \dots, w_r of_\ell$ . We denote the  $j$ th entry of  $sl_i$  by  $sl_{j,i}$ .

2. Obtain the first  $\ell$  predictions from our recursive forecasting model as follows:

Step 1. Make a prediction on observation 1 from the testing set using  $w_r of_0$ . Call this prediction  $op_1$ .

Step 2. Set  $sl_{2,1}$  equal to  $op_1$ . Make a prediction on observation 2 from the testing set and  $sl_{2,1}$ , using  $w_r of_1$ . Call this prediction  $op_2$ .

Step 3. Set  $sl_{3,1}$  equal to  $op_2$ , and set  $sl_{3,2} = sl_{2,1}$ . Make a prediction on observation 3, from the testing set and  $sl_{3,1}, sl_{3,2}$  using  $w_r of_2$ . Call this prediction  $op_3$ .

$\vdots \quad \vdots \quad \vdots \quad \vdots$

Step  $\ell$ . Set  $sl_{\ell,1}$  equal to  $op_{\ell-1}$  and set

$$sl_{\ell,2} = sl_{\ell-1,1}, \dots, sl_{\ell,\ell-1} = sl_{\ell-1,\ell-2}.$$

Make a prediction on observation  $\ell$  from our testing set and  $sl_{1,\ell}, \dots, sl_{\ell,\ell-1}$  using  $w_r of_{\ell-1}$ . Call this prediction  $op_\ell$ .

3. For  $g = \ell + 1$  to  $n$ , set  $sl_{g,1} = op_{g-1}$  and by set

$$sl_{g,2} = sl_{g-1,1}, \dots, sl_{g,\ell} = sl_{g-1,\ell-1}.$$

Make a prediction on observation  $g$  of our testing set and  $sl_{g,1}, \dots, sl_{g,\ell}$  using  $w_r of_\ell$ . Call this prediction  $op_g$ .

In Table 5, we demonstrate the recursive forecasting process with  $\ell = 4$  lagging periods. For each observation in the testing set, one of  $w_r of_k$ , for  $k = 1, \dots, \ell$ , makes a prediction from the values of the explanatory variables in the corresponding row. These explanatory variables are accelerometer information and values of the “psudeo-lagging” variables which are prior predictions. Models  $w_r of_k$  for  $k = 1, \dots, \ell$  are specified in section 4.3.3.

### 4.3.3 Training model weights

We first fit a full ordinal forest model, with accelerometer information and demographic variables, to all the training data. As explained in section 4.2, this returns final prediction intervals and their midpoints,  $m_1, \dots, m_c$ , where  $c$  is the number of ordinal classes ( $c = 4$  in our data). The ordinal

Table 5: Recursive Prediction for Participant  $r$  in Testing with  $\ell = 4$ 

Obs	$x_1$	$x_2$	$\dots$	$x_p$	$sl_1$	$sl_2$	$sl_3$	$sl_4$	Model	Prediction
1	$x_{11}$	$x_{12}$	$\dots$	$x_{1p}$					$w_r of_0$	$op_1$
2	$x_{21}$	$x_{22}$	$\dots$	$x_{2p}$	$op_1$				$w_r of_1$	$op_2$
3	$x_{31}$	$x_{32}$	$\dots$	$x_{3p}$	$op_2$	$op_1$			$w_r of_2$	$op_3$
4	$x_{41}$	$x_{42}$	$\dots$	$x_{4p}$	$op_3$	$op_2$	$op_1$		$w_r of_3$	$op_4$
5	$x_{51}$	$x_{52}$	$\dots$	$x_{5p}$	$op_4$	$op_3$	$op_2$	$op_1$	$w_r of_4$	$op_5$
6	$x_{61}$	$x_{62}$	$\dots$	$x_{6p}$	$op_5$	$op_4$	$op_3$	$op_2$	$w_r of_4$	$op_6$
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$n$	$x_{n1}$	$x_{n2}$	$\dots$	$x_{np}$	$op_{n-1}$	$op_{n-2}$	$op_{n-3}$	$op_{n-4}$	$w_r of_4$	$op_n$

forest package (Hornung, 2019) also returns variable importance measures for all of the explanatory variables. Let  $v_1, \dots, v_d$  be these measures for our four time-invariant demographic variables.

We now compute the sample covariance matrix,  $S$ , from the time-invariant demographic variables of the training participants. These variables include the variables age, weight, height and sex (coded as 0 or 1). For participant  $r$  in the testing set, we compute the weighted Mahalanobis distances (Kotu and Deshpande, 2018) between the values of this participant's demographic information and each of the other  $T$  training participants' demographic information using  $v_1, \dots, v_d$  as weights. That is, with diagonal matrix  $W = \text{diag}(v_1, \dots, v_d)$ , demographic values  $x_r = (x_{r,1}, \dots, x_{r,d})$  for participant  $r$  and with  $x_h = (x_{h,1}, \dots, x_{h,d})$  for training participant  $h$ , we compute weighted Mahalanobis distance

$$f_{r,h} = \sqrt{(x_r - x_h) W S^{-1} (x_r - x_h)'}$$

Then we set

$$w_{r,h} = \frac{(f_{r,h} + \epsilon)^{-1}}{\sum_{j=1}^T (f_{r,j} + \epsilon)^{-1}} \text{ for } h = 1, \dots, T \text{ with } \epsilon > 0. \quad (4.1)$$

The addition of  $\epsilon > 0$  avoids overly large weights when the demographic information of testing participant  $r$  is very similar to the demographic information of training participant  $h$ . A smaller  $\epsilon$  will give larger relative weights to similar participants, while a larger  $\epsilon$  will incorporate weights and information from more participants. In 3. of our recursive forecasting model of section 4.3.2, to obtain a prediction from  $w_r of_\ell$ , using observation  $g$  from testing participant  $r$  and values of pseudo-lagging variables  $s_{g,1}, \dots, s_{g,\ell}$ , we first obtain predictions from  $of_{1,\ell}, \dots, of_{T,\ell}$ , according with the mid-points,  $m_1, \dots, m_c$ , of prediction intervals. Denoting these predictions  $p_{1,\ell}, \dots, p_{T,\ell}$  we take the weighted average

$$A_{r,\ell} = \sum_{h=1}^T w_{r,h} \cdot p_{h,\ell}$$

The prediction from  $w_r of_\ell$  is the the class associated with midpoint

$$\operatorname{argmin}_{j=1,\dots,c} |A_{r,\ell} - m_j|.$$

Similarly, in the steps of 2. of the recursive forecasting model in section 4.3.2, to obtain a prediction from  $w_r of_k$ , we take a weighted average of predictions from  $of_{1,k}, \dots, of_{T,k}$ , with the weights computed from the weighted Mahalanobis distances to the demographic information to each participant in training set. These weights account for both the similarity between the testing participants and the training participants and the relative importance of the demographic variables in determining activity level patterns. In Figure 2 we have a schematic that summarizes our recursive forecast model with multiple time series.

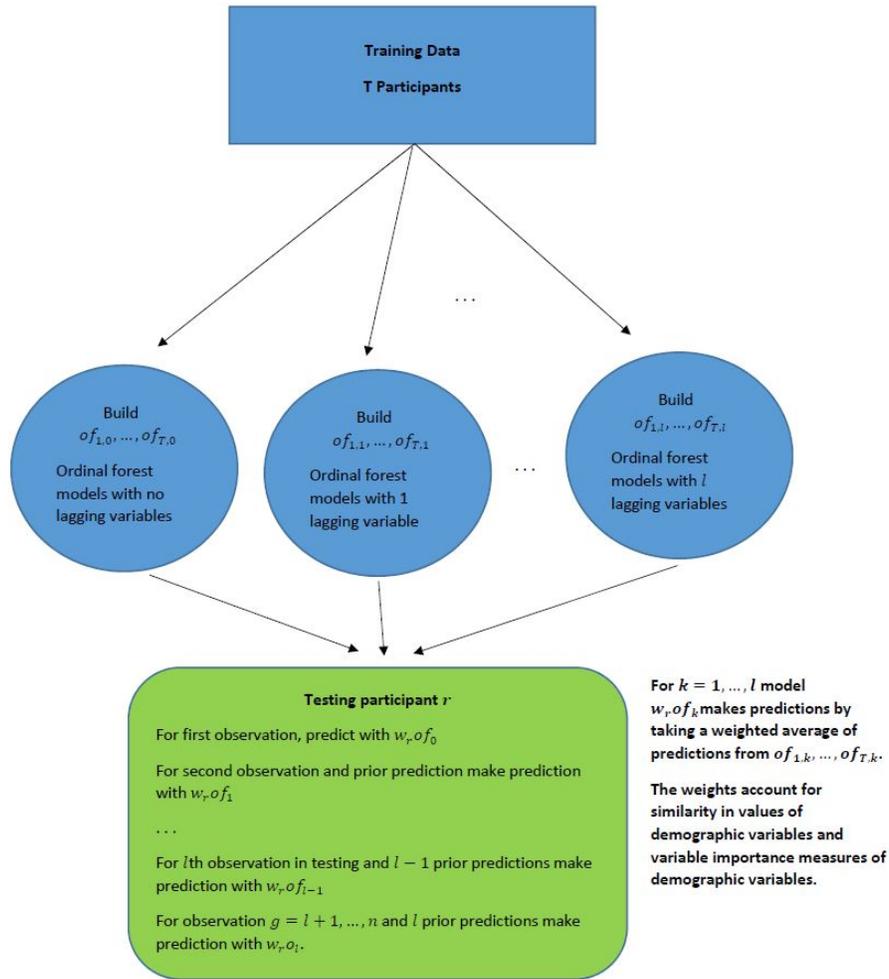


Figure 2: Recursive Forecast with Multiple Time Series

## 4.4 Measures of Performance

We assess the performance of our models with testing error rates, with weighted kappa values and the ordinal gamma statistic.

### 4.4.1 Testing Error Rates

The error rate on the testing set is simply the proportion of predictions that the model classifies incorrectly according to the actual observations of the response variable in the test set. The testing error rate does not take into account the ordinality of the responses. For example, a classification of SED as LPA is just as significant as a misclassification of SED as VPA even though in the former case the prediction is “one away” from the correct observation and in latter case the prediction is “three away” according to the order of the responses.

### 4.4.2 Kappa Statistic

As presented in Hossain et al. (2021), kappa values compare observed versus expected accuracies to account for classification by chance Landis and Koch (1977). We look at linearly weighted kappa values for four class classifications. Weighted kappa values treat misclassifications further away from true classifications as more significant than misclassifications closer to true classifications Cohen (1968). For example, we want to penalize a misclassification of SED as light LPA as less significant than a misclassification of sedentary SED as MPA or VPA. We use a linearly weighted kappa measure to do so, with movement from the true classification from one class to another penalized equally. A scale given in Landis and Koch (1977) and often used to interpret kappa values is presented in Table 6.

Table 6: Interpreting Kappa

Agreement	Slight	Fair	Good	Substantial	Almost Perfect
Kappa Value	0.01–0.20	0.21–0.40	0.41–0.60	0.61–0.80	0.81–0.99

### 4.4.3 Ordinal Gamma Statistic

The ordinal gamma statistic as formulated in Agresti (2003), is another assesment of the predictive accuracy of a decision that does take into account the order of the responses. It considers the number of concordant and discordant pairs from the confusion matrix in weighted sums, with weights agreeing with the order of the data. The ratio of the difference of these sums from concordant pairs,  $C$ , and discordant pairs,  $D$  is then taken to compute the gamma statistic as

$$\gamma = \frac{C - D}{C + D}.$$

This measure,  $\gamma$ , is between  $-1$  and  $1$  with a greater magnitude indicating a stronger association. Negative numbers indicate negative association according to the order of the data (for example, if SED was often predicted as MPA or VPA), and positive numbers indicate a positive association according to the order of the data (for example, if SED was often predicted as SED or LPA).

## 5 Results

In Tables 7, 8, 9, 10 we assess our random forest, ordinal forest and recursive forecasting models (with  $\ell = 1, 2$  and  $3$  lag periods) built on our training data with assessments from section 4.4. The models are trained on the data from the participants from the first visit and tested on data from the participants in the second visit, as specified in section 3.3. The recursive forecasting models use time series information in training, and weighted, prior predictions with ordinal forest models in testing, as specified in section 4.3.

Table 7: Assessment of Models on Ankle Placement

Model	Assessment Measure		
	Error Rate	W. Kappa	Ord. Gamma
Random Forest	0.337	0.647	0.868
Ordinal Forest	0.329	0.661	0.881
Rec. Forecast ( $\ell = 1$ )	0.286	0.689	0.901
Rec. Forecast ( $\ell = 2$ )	0.280	0.701	0.910
Rec. Forecast ( $\ell = 3$ )	0.321	0.659	0.891

Table 8: Assessment of Models on Left Wrist Placement

Model	Assessment Measure		
	Error Rate	W. Kappa	Ord. Gamma
Random Forest	0.351	0.610	0.842
Ordinal Forest	0.340	0.631	0.851
Rec. Forecast ( $\ell = 1$ )	0.290	0.695	0.897
Rec. Forecast ( $\ell = 2$ )	0.287	0.701	0.899
Rec. Forecast ( $\ell = 3$ )	0.325	0.672	0.878

In Figure 3, we see an agreement plot (Bangdiwala and Shankar, 2013) for the recursive forecast model with  $\ell = 1$  and the ankle placement. As in Hossain et al. (2021), the width of the black

Table 9: Assessment of Models on Right Wrist Placement

Model	Assessment Measure		
	Error Rate	W. Kappa	Ord. Gamma
Random Forest	0.373	0.573	0.820
Ordinal Forest	0.369	0.596	0.839
Rec. Forecast ( $\ell = 1$ )	0.302	0.644	0.857
Rec. Forecast ( $\ell = 2$ )	0.305	0.652	0.861
Rec. Forecast ( $\ell = 3$ )	0.371	0.582	.819

Table 10: Assessment of Models on Hip Placement

Model	Assessment Measure		
	Error Rate	W. Kappa	Ord. Gamma
Random Forest	0.365	0.609	0.828
Ordinal Forest	0.360	0.631	0.857
Rec. Forecast ( $\ell = 1$ )	0.295	0.691	0.869
Rec. Forecast ( $\ell = 2$ )	0.289	0.695	0.886
Rec. Forecast ( $\ell = 3$ )	0.337	0.659	0.870

inner squares in each rectangle is the number of agreements for each particular class and the gray shading represents misclassifications of adjacent classes. The plot illustrates that the number of misclassifications for any class outside of classes adjacent to that class, is relatively small. This is reflected in the substantial agreement in the linearly weighted kappa values which account for ordinality in classification.

## 6 Discussion

In this paper, we incorporated the ordinality and time series nature of accelerometer data, with responses in four-activity level classes into a recursive, multiple time series forecasting model. We used tree-based models, and ordinal forest, which takes into account the ordinality of responses, in assigning quantitative values to the responses. A forecasting model was developed which makes recursive predictions based on current raw, accelerometer summaries and on prior predictions. These predictions take into account similarity of participants in the training set to participants in the testing set in terms of demographic variables and measures of variable importance through the weighted Mahalanobis distance. In our results, ordinal forest shows little difference with random forest in er-

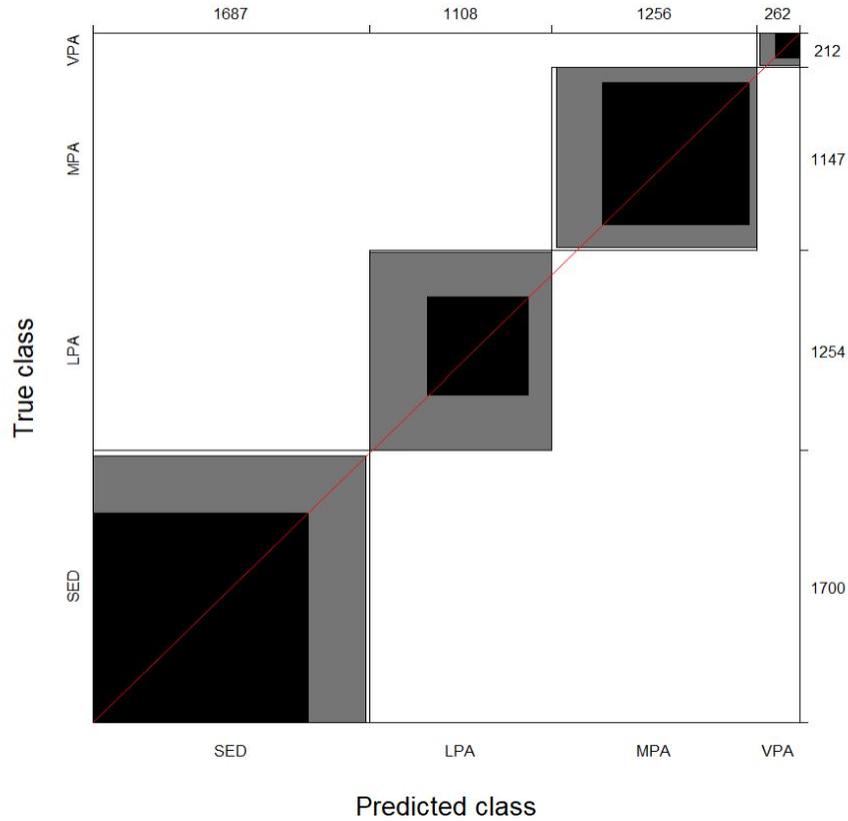


Figure 3: Agreement Plot: Recursive Forecasting ( $\ell = 1$ )

ror rates and slight improvements in linearly weighted kappa values. This agrees with Hornung (2019) and can be improved by incorporating a weighted splitting criteria in random forest such as the weighted Gini index or the weighted entropy that takes account ordinality in assessing impurity of splits. The recursive forecasting model shows substantial improvement in all assessment measures for  $\ell = 1$  and  $\ell = 2$ . However, including additional lagging periods with  $\ell$  greater than 2, although providing additional information about activity patterns, causes the model to perform worse in testing. This can be investigated and be the subject of additional model development. Model weights and parameters such as  $\epsilon$  in (4.1) in section 4.3.3 can be adjusted by validation in training. Other types of ordinal regression models, such as proportional-odds and neural network models (Cao et al., 2020), can be incorporated into the framework provided in this paper. This approach can be applied to financial, physiological and medical data, among others, which have multiple time series with ordinal responses.

## References

- ActiGraphCorp (2017), “ActiGraph White Paper: What is A Count?” Tech. rep.
- Agresti, A. (2003), *Categorical data analysis*, vol. 482, John Wiley & Sons.
- Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett Jr, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., and Leon, A. S. (2011), “2011 Compendium of Physical Activities: a second update of codes and MET values,” *Medicine & science in sports & exercise*, 43, 1575–1581.
- Amato, F., Jacques, J., and Prim-Allaz, I. (2024), “Clustering longitudinal ordinal data via finite mixture of matrix-variate distributions,” *Statistics and Computing*, 34, 81.
- Bangdiwala, S. I. and Shankar, V. (2013), “The agreement chart,” *BMC medical research methodology*, 13, 1–7.
- Branch, W. A. and Evans, G. W. (2006), “A simple recursive forecasting model,” *Economics Letters*, 91, 158–166.
- Brant, R. (1990), “Assessing proportionality in the proportional odds model for ordinal logistic regression,” *Biometrics*, 1171–1178.
- Breiman, L. (2001), “Random forests,” *Machine learning*, 45, 5–32.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984), *Classification and regression trees*, CRC press.
- Brieman, L., Cutler, A., Liaw, A., and Weiner, M. (2018), “Package ‘randomforest’,” *University of California, Berkeley: Berkeley, CA, USA*.
- Cao, W., Mirjalili, V., and Raschka, S. (2020), “Rank consistent ordinal regression for neural networks with application to age estimation,” *Pattern Recognition Letters*, 140, 325–331.
- Caspersen, C. J., Pereira, M. A., and Curran, K. M. (2000), “Changes in physical activity patterns in the United States, by sex and cross-sectional age,” *Medicine & Science in Sports & Exercise*, 32, 1601–1609.
- Cohen, J. (1968), “Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit,” *Psychological bulletin*, 70, 213.
- De Vries, S. I., Garre, F. G., Engbers, L. H., Hildebrandt, V. H., and Van Buuren, S. (2011), “Evaluation of neural networks to identify types of activity using accelerometers,” *Med Sci Sports Exerc*, 43, 101–7.
- Ellis, K., Kerr, J., Godbole, S., Lanckriet, G., Wing, D., and Marshall, S. (2014), “A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers,” *Physiological measurement*, 35, 2191.

- Hornung, R. (2019), "Ordinal forests," *Journal of Classification*, 1–14.
- Hossain, S. S., Lazar, D. M., and Begum, M. (2021), "Ordinal Statistical Models of Physical Activity Levels from Accelerometer Data," *International Journal of Exercise Science*, 14, 338–357.
- Janitza, S., Tutz, G., and Boulesteix, A.-L. (2016), "Random forest for ordinal responses: prediction and variable selection," *Computational Statistics & Data Analysis*, 96, 57–73.
- Kotu, V. and Deshpande, B. (2018), *Data science: concepts and practice*, Morgan Kaufmann.
- Landis, J. R. and Koch, G. G. (1977), "The measurement of observer agreement for categorical data," *biometrics*, 159–174.
- Lazar, D., Begum, M., Murshed, M. M., Nelson, B., Bock, J. M., Imboden, M., Kaminsky, L., and Montoye, A. (2020), "Statistical Learning Methods to Predict Activity Intensity from Body-Worn Accelerometers," *Journal of Biomedical Analytics*, 3, 27–50.
- Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Watanabe, Y. Y., Murgatroyd, M., and Papastamatiou, Y. P. (2017), "Analysis of animal accelerometer data using hidden Markov models," *Methods in Ecology and Evolution*, 8, 161–173.
- Montoye, A. H., Begum, M., Henning, Z., and Pfeiffer, K. A. (2017a), "Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data," *Physiological measurement*, 38, 343.
- Montoye, A. H., Nelson, M. B., Bock, J. M., Imboden, M. T., Kaminsky, L. A., Mackintosh, K. A., McNarry, M. A., and Pfeiffer, K. A. (2017b), "Raw and count data comparability of hip-worn ActiGraph GT3X+ and link accelerometers," *Medicine & Science in Sports & Exercise*, 50, 1103.
- Montoye, A. H., Westgate, B. S., Fonley, M. R., and Pfeiffer, K. A. (2018), "Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer," *Journal of Applied Physiology*, 124, 1284–1293.
- Nawaratne, R., Alahakoon, D., De Silva, D., O'Halloran, P. D., Montoye, A. H., Staley, K., Nicholson, M., and Kingsley, M. I. (2020), "Deep Learning to Predict Energy Expenditure and Activity Intensity in Free Living Conditions using Wrist-specific Accelerometry," *Journal of Sports Sciences*, 1–8.
- Ng, C. N. and Young, P. C. (1990), "Recursive estimation and forecasting of non-stationary time series," *Journal of Forecasting*, 9, 173–204.
- Nickel, C., Busch, C., Rangarajan, S., and Möbius, M. (2011), "Using hidden markov models for accelerometer-based biometric gait recognition," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, IEEE, pp. 58–63.
- Pavey, T. G., Gilson, N. D., Gomersall, S. R., Clark, B., and Trost, S. G. (2017), "Field evaluation of a random forest activity classifier for wrist-worn accelerometer data," *Journal of science and medicine in sport*, 20, 75–80.

- Robert, B., White, B., Renter, D., and Larson, R. (2009), "Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle," *Computers and Electronics in Agriculture*, 67, 80–84.
- Sallis, J. F. (2000), "Age-related decline in physical activity: a synthesis of human and animal studies," *Medicine & Science in Sports & Exercise*, 32, 1598–1600.
- Sasaki, J. E., Hickey, A., Staudenmayer, J., John, D., Kent, J. A., and Freedson, P. S. (2016), "Performance of activity classification algorithms in free-living older adults," *Medicine and science in sports and exercise*, 48, 941.
- Statistics, B. L. (2008), "American time use survey," Retrieved June, 16, 2020.
- Taieb, S. B., Hyndman, R. J., et al. (2012), *Recursive and direct multi-step forecasting: the best of both worlds*, vol. 19, Citeseer.
- Tan, H. X., Aung, N. N., Tian, J., Chua, M. C. H., and Yang, Y. O. (2019), "Time series classification using a modified LSTM approach from accelerometer-based data: A comparative study for gait cycle detection," *Gait & posture*, 74, 128–134.
- Tangirala, S. (2020), "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm," *International Journal of Advanced Computer Science and Applications*, 11, 612–619.
- Troiano, R. P., McClain, J. J., Brychta, R. J., and Chen, K. Y. (2014), "Evolution of accelerometer methods for physical activity research," *British journal of sports medicine*, 48, 1019–1023.
- Tyralis, H. and Papacharalampous, G. (2017), "Variable selection in time series forecasting using random forests," *Algorithms*, 10, 114.
- Ward, D. S., Evenson, K. R., Vaughn, A., Rodgers, A. B., and Troiano, R. P. (2005), "Accelerometer use in physical activity: best practices and research recommendations," *Medicine and science in sports and exercise*, 37, S582–8.
- Witowski, V., Foraita, R., Pitsiladis, Y., Pigeot, I., and Wirsik, N. (2014), "Using hidden markov models to improve quantifying physical activity in accelerometer data—a simulation study," *PloS one*, 9, e114089.

Received: February 26, 2024

Accepted: April 3, 2024