

A Functional Data Analysis of Accelerometer-Based Activity

by

Chandler Capps

BS, Mathematics, BA, Economics, University of Georgia, 2018, 2018

Submitted to the Graduate Faculty of the

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Chandler Capps

It was defended on

June 12, 2020

and approved by

Thesis Advisor:

Robert T. Krafty, PhD

Associate Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Chaeryon Kang, PhD

Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Stephen F. Smagula, PhD

Assistant Professor, Department of Psychiatry
University of Pittsburgh

Copyright © by Chandler Capps

2020

A Functional Data Analysis of Accelerometer-Based Activity

Chandler Capps, MS

University of Pittsburgh, 2020

Abstract

Introduction: Depression is prevalent, often comorbid with other negative health outcomes, and can have severe direct and indirect consequences. It is critical, then, that research contributes to the identification and prediction of depression cases. Accordingly, scores of studies have shown statistically significant associations between certain rest-activity rhythm (RAR) characteristics and current presence of depression symptoms and future depression development. However, the study of age's moderation of the effects of the presence of depression symptoms on RARs has yet to be thoroughly explored. This study attempts to provide insight into the interaction between age and depression status in predicting binned log activity counts and to offer a unique exploratory analysis via functional methods.

Methods: This study focuses on functional linear modelling, in particular, functional ANOVA (fANOVA). The model's response variable is a vector of functions constructed through the interpolation and penalized smoothing of the data. These functions map time to mean log activity count. The predictors are di- or polychotomous and include depression status, age group, gender, and ethnicity. fANOVA yields least-squares estimates of the functional effects of each covariate on the outcome. A permutation F-test determines whether the full model, including depression and age group interactions, explains significantly more variation in the data than the main-effects model.

Results: The fANOVA yielded statistically significant results at a .05 significance level for all covariates. In particular, depression's negative effect on mean log activity count is statistically significant and pronounced for all waking hours (approximately 6:00 am to 10:00 pm). The young adulthood effect is similar while the pronounced effects of later adulthood and older adulthood are positive. The sole statistically significant interaction effect, depression and young adulthood, is positive and mitigates much of the effect of depression. The permutation F-test determined the full model to be statistically significant at a .05 significance level.

Conclusion: The evidence suggests that depression is statistically significantly associated with activity over typical waking hours, and the effect is moderated by age in that young adults experience a considerably lessened depression effect.

Public Health Significance: This study contributes to the understanding of the observable manifestation of depression in adults.

Table of Contents

1.0 Introduction.....	1
1.1 RARs and Accelerometry	1
1.2 RAR Methodology	2
1.3 The Present Study.....	4
2.0 Data	5
2.1 Preliminary Data	5
2.1.1 Data Origin.....	5
2.1.2 Accelerometer and Protocol.....	5
2.2 Model Data	6
2.3 Outcome.....	6
2.4 Covariates.....	6
3.0 Statistical Methods.....	8
3.1 Exploratory Analysis.....	8
3.2 Functional Analysis of Variance	11
4.0 Results	13
4.1 Exploratory Results.....	13
4.2 Functional Analysis of Variance Results	19
4.2.1 Estimated Coefficient Functions.....	20
4.2.2 Permutation F-test Results.....	22
5.0 Conclusions.....	23
Appendix R Code	24

Bibliography 31

List of Tables

Table 1 Percentage of subjects in each group who are classified as depressed.....	7
--	----------

List of Figures

Figure 1 Standard five-parameter extended cosine curve.	3
Figure 2 Generalized cross-validation criterion as a function of $\log(\lambda)$.	10
Figure 3 Ten smoothed mean log activity count curves from data subsets by depression status.....	14
Figure 4 Estimated autocovariance surfaces for the mean log activity count data.	15
Figure 5 Contour plots of the bivariate variance-covariance surfaces by hours since midnight.....	16
Figure 6 A phase-plane plot for one subject.....	18
Figure 7 Estimated regression coefficients for predicting log activity count.	20
Figure 8 Permutation F-statistics Plot.	22

1.0 Introduction

1.1 RARs and Accelerometry

Today, there is compelling evidence that mood, memory, cognition, and other aspects of neurobehavioral and neurocognitive functioning are associated with characteristics of our rest-activity rhythms [10]. Rest-activity rhythms, or RARs, can be described as individual, daily activity profiles with several characteristics such as shape, scale, timing, and fit to a 24-hour pattern. It is these attributes that empirically driven models of RARs seek to emulate through the estimation of a model-determined set of parameters that ultimately informs the more general attributes. Similarly, deterministic measures of RARs are also used to create an empirical, object account of RARs.

The objectivity of these accounts is achieved through the use of an accelerometer. A triaxial accelerometer is a small device that measures acceleration in three directions; this data can then be integrated into a single activity count representing overall movement. This unobtrusive device eliminates the avenues for potential recall and self-report bias that exist in other activity reporting methods such as sleep journals.

Given the vantage point that accelerometry data provides, dated and recent studies alike have adopted its use in investigating possible associations between RAR characteristics and neurocognitive and neurobehavioral functioning. As previously suggested, many of these investigations yielded statistically significant results. In fact, statistically significant results can be found in both cross-sectional and longitudinal studies. For example, a survey of recent cross-sectional studies showed that there is consistent evidence that RARs characterized by less stability

[11], greater fragmentation [3, 7, 8, 11], lower amplitude, and worse fit [7, 19] were found to be statistically significantly associated with dementia diagnosis. Furthermore, future cognitive impairment was shown to be associated with RAR measures, specifically fit and lower amplitude [1, 22]. Associations between depression symptom presence and a wide range of RAR characteristics also exist in relevant literature. Though the significance of such associations was often attenuated with the addition of other basic health or lifestyle factors, statistically significant associations with low RAR robustness and low mesor tended to persist [9, 11-13, 16, 17]. Moreover, like future cognitive impairment, RAR measures have proven to be useful in the prediction of future depression development. In particular, low RAR robustness was statistically significant in association with twice higher odds of future depression symptoms [21].

1.2 RAR Methodology

It is conventional to differentiate between two historically predominant methods of analyzing accelerometry data: parametric and nonparametric. Within the domain of parametric approaches, the truncated Fourier series and the cosinor models (along with its extensions and transformations) dominate. The latter enjoys particular popularity due to the high interpretability of its parameters and its relative parsimony. For example, the sigmoidally transformed extended cosinor models provide an intuitive account of the RAR with five parameters estimated by nonlinear least squares estimation: α , β , ϕ , amplitude, and mesor. This class of models also offers a sort of goodness of fit measure, the pseudo-F statistic. However, a disadvantage of individual cosinor models is that the validity of the model relies upon the validity of the class of cosinor models. Nonparametric methods offer freedom from model assumptions and offer interpretable,

informative RAR measures. They provide deterministic measures such as interdaily stability (IS), intradaily variability (IV), L5, and M10, and RA. IS measures the strength of the circadian rhythm, and IV quantifies the fragmentation in the rhythm. Thus, both IS and IV are measures of variability in the RAR and, in that sense, can be thought of as deterministic analogues to the pseudo-F statistic of the cosinor models. L5 and M10, respectively, represent the mean activity of the five-hour period with the least total activity and the mean activity of the ten-hour period with the highest total activity. RA is, then, the relative amplitude of the two periods, L5 and M10.

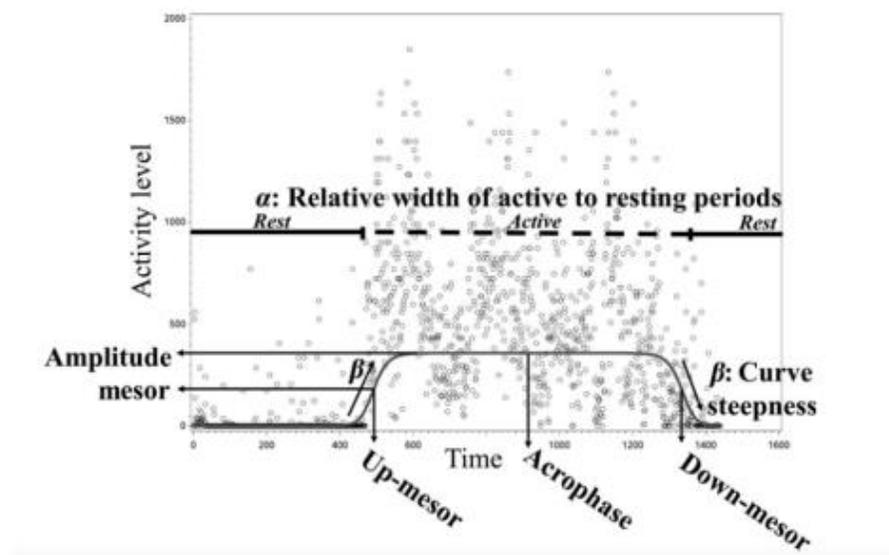


Figure 1 Standard five-parameter extended cosine curve.

Credit: Smagula et al. (2007)

In the recent past, studies have also used functional regression in analyzing RAR characteristics [6, 25]. A functional regression approach uses a set of basis functions and corresponding estimated coefficient functions to construct a mean functional profile. Two relevant studies reported significant differences in RAR characteristics by age [6, 25]. In particular, both studies reported not only that the profiles of older adults tend to show lower activity while controlling for gender or BMI, but that the effect was most pronounced in the late afternoon. Xiao

et al goes further in reporting that the duration of active periods, amount of activity, amount of moderate/vigorous activity, and activity peaks are all decreasing as a result of aging. Moreover, the late-afternoon age effect is larger for women, and there is evidence to suggest that women tend to have longer periods of light activity while men engage in short periods of intense activity [25].

1.3 The Present Study

The present study aims to contribute to the understanding of the relationship between depression status and RARs. In particular, we hope to ascertain the significance of depression status as predictor of mean log activity count for specific times throughout the day. Depressed individuals are thought to present differences in RARs from non-depressed individuals such as different timing in rise to peak activity, lower activity over waking hours, and increased activity during non-waking hours [21]. These effects hint at possible statistically significant differences between groups. Also, we hope to shed light on how or if age moderates depression's effects on log activity count. Secondly, this paper is meant to demonstrate fANOVA as a useful tool in the analysis of RARs. The quasi-sinusoidal nature of RARs makes them an appropriate application for Fourier-based functional data analysis. Moreover, using functional data analysis will allow us to provide quality visual insight into variance differences between groups and estimated covariate effects as functions of time.

2.0 Data

2.1 Preliminary Data

2.1.1 Data Origin

Both the physical activity data and the demographics data were originally sourced from the Centers for Disease Control's health assessment program, NHANES, or the National Health and Nutrition Examination Survey. The data was gathered via examination and questionnaire administered to subjects selected by thirty survey locations across the United States between 2005 and 2006.

2.1.2 Accelerometer and Protocol

According to the NHANES actigraphy data codebook, the device used in NHANES was the ActiGraph AM-7164 (formerly the CSA/MTI AM-7164), manufactured by ActiGraph of Ft. Walton Beach, FL. As mentioned, the device is programmed to detect and record the magnitude of acceleration or intensity of movement acceleration. A one-minute epoch was used in NHANES. Records were summed over each one-minute epoch.

Subjects were recruited for the mobile examination center component. A health technician explained the protocol to potential subjects. The monitors were programmed to begin recording activity information beginning at 12:01 a.m. the day following the examination of the corresponding subject. Subjects wore the device on the right hip in a fitted, elastic belt. Protocol dictated that the device remain dry and be removed at bedtime [3, 4].

2.2 Model Data

Both the physical activity and demographics datasets were thinned separately to include only complete cases of relevant variables. After thinning, the physical activity dataset was aggregated such that each subject would account for 24 observations, one for each hour time bin. Each observation contained the mean $\log(\text{activity count} + 1)$, where mean aggregates over all minute-level data across all seven days. The datasets were thinned again leaving 4391 subjects with complete physical activity and demographic data.

2.3 Outcome

The outcome of interest in the present study is a function for which 24-hour time is mapped to an estimated log mean activity count.

2.4 Covariates

Primary among the covariates of this study is depression status. This variable is binary and reflects a PHQ-9 score with a threshold of 10. Age group is included as another covariate of interest and as a possible moderator of depression's effect on mean log activity count. Age group is defined as a polychotomous variable with four levels: ages 0-30, corresponding to young adulthood; 31-45, adulthood; 46-60, later adulthood; 61+, older adulthood. Other covariates were considered which have relevance to the generalizability of our findings. In particular, we consider self-

reported gender and ethnicity. These demographics are defined by the categories “Male” and “Female,” and “Mexican,” “Other Hispanic,” “White,” “Black,” and “Other”, respectively.

Table 1 Percentage of subjects in each group who are classified as depressed.

	Age Group			
	Young Adulthood	Adulthood	Later Adulthood	Older Adulthood
Total	6	6	9	4
Ethnicity				
White	6	6	9	4
Black	6	5	20	< 1
Mexican	4	5	8	3
Other Hispanic	8	6	12	7
Other	2	7	13	5
Gender				
Male	4	5	8	4
Female	7	7	11	4

3.0 Statistical Methods

3.1 Exploratory Analysis

Functional data analysis enjoys a unique advantage in accomplishing common goals of exploratory analyses. Through interpolation, smoothing, and derivation, functional data analysis offers opportunities to more effectively highlight important data characteristics and identify patterns of variation. In the present study we use penalized smoothing techniques to construct mean log activity count functions of time. We use a bivariate covariance function to visually identify autocovariance features in RARs, and we examine the relationship between the velocity and acceleration of activity throughout the 24-hour period for one subject.

As suggested above, smoothing is an integral part of any functional data analysis. It is the workhorse by which the functional data analyst can construct approximations of functions that underlie the data, and it is where we begin. We construct the estimated mean log activity count functions by defining a set of foundational components: the basis functions. Then we estimate a matrix of coefficients to linearly combine the basis functions in such a way as to yield the estimated mean log activity count function. Each resulting function is of the following form:

$$y_i(t) = \sum_{k=1}^K c_{i,k} \phi_k(t)$$

where $i \in [1, I]$ where $I=4391$, the number of subjects; and $k \in [1, K]$ where $K=31$, the number of basis functions.

We begin by choosing to work in the Fourier basis system due to the ease of computation, periodicity, agreement with the shape of RAR data, and the lack of need for greater flexibility that

can be obtained via B-splines. We define a measure of roughness as the harmonic acceleration operator.

The roughness penalty essentially limits the departure of the curve from some smooth baseline and protects against overfitting. The harmonic acceleration operator was chosen because when applied to the sum of the first three terms of the Fourier series, it yields zero; and that yield only increases in cubes of highest coefficient of ωt as more sine and cosine pairs are added. Because we are using a roughness penalty in estimating our activity profiles, it is typical to select a number of basis functions slightly greater than the number of observations per subject. In this case, we choose to use 31 Fourier basis functions. We now have an error function, the minimization of which will yield coefficient estimates and ultimately estimated activity profiles. The error function for a particular coefficient vector is given below:

$$F(c) = \sum_l [q_l - y(t_l)]^2 + \lambda \int [Lx(t)]^2 dt$$

where $l \in [1, L]$ where $L=24$, the number of observations per subject; and λ determines the weight of penalization.

The value of λ is chosen using the generalized cross-validation (GCV) criterion proposed by Craven and Wahba [5]. The criterion is a multiple of an unbiased error variance estimate. Accordingly, we choose λ through minimization of the $GCV(\lambda)$.

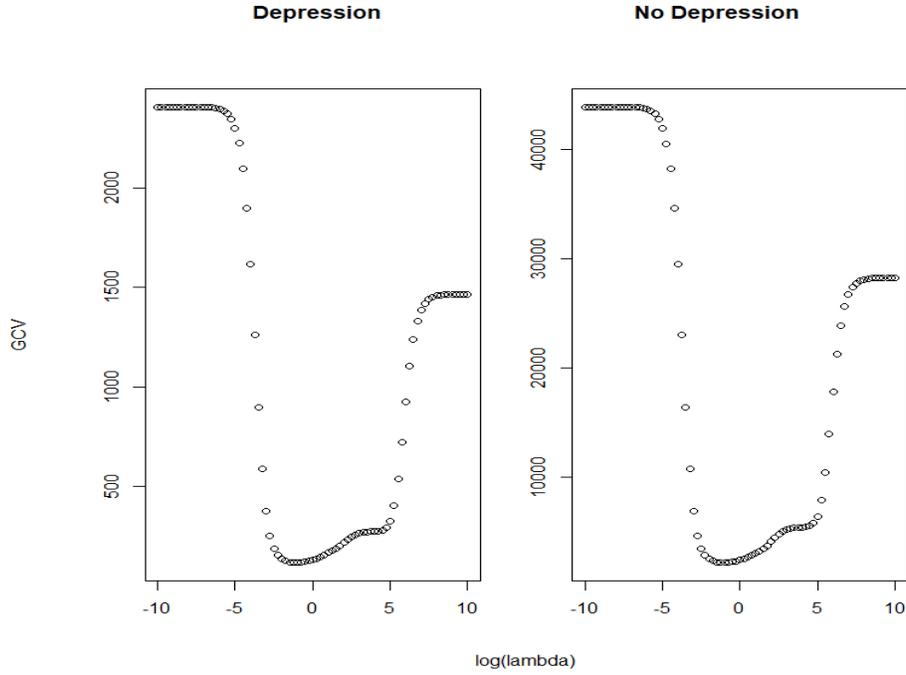


Figure 2 Generalized cross-validation criterion as a function of $\log(\lambda)$.

As mentioned, after a functional data object is created through least squares estimation with roughness penalization, we proceed to examine the relationship between derivatives of the fitted curve and to provide some descriptive measures. The phase-plane plot displays the first derivative of the function against the second, or velocity and acceleration. The plot is often described in terms of the variation of the potential and kinetic energy in a system. We too take this approach. With regard to the bivariate covariance function, it is necessary to specify that the function yields the covariance between curve values $y_i(s)$ and $y_i(t)$ where s and t are times and i indicates the estimated mean log activity count function for the i th subject. The corresponding equation is given below.

$$cov(s, t) = \frac{1}{N-1} \sum_i [y_i(s) - \bar{y}(s)][y_i(t) - \bar{y}(t)]$$

where N is the total number of observations.

3.2 Functional Analysis of Variance

Here we aim to describe the primary methodology of the present study: the functional analysis of variance, or fANOVA. fANOVA models the relationship between a linear combination of scalar inputs and a functional response by decomposing the functional response into functional covariate effects. In the case of the present study, we have a model of the form:

$$y_i(t) = \beta_0(t) + \sum_j^{J-1} x_{ij} \beta_j(t) + \varepsilon_i(t)$$

where j indicates a particular level of a covariate, a dummy variable.

In this case, the values of x_{ij} are either 0 or 1, and only one nonzero instance of x_{ij} can exist over all dummy variables belonging to a single covariate for each individual. Furthermore, it is necessary that the effects be constrained as illustrated below in order to uniquely identify the effects.

- $\forall m$ covariates corresponding to main effects $\sum_j \beta_{m_j}(t) = 0$ where $m_j \in [1, M_j]$ and M_j is the number of levels of main effect m
- For all levels of age group, $\beta_{depression,age\ group}(t) + \beta_{no\ depression,age\ group}(t) = 0$
- $\beta_{depression,young\ adulthood}(t) + \beta_{depression,adulthood}(t) + \beta_{depression,late\ adulthood}(t) + \beta_{depression,older\ adulthood}(t) = 0$

To estimate the functional covariate effects, we again turn to ordinary least squares. We just have to switch the usual discrete summation for integration. Furthermore, we would still like a measure of the precision with which we have calculated the coefficients of the functional effects. Fortunately, statistical packages in R have made this accessible without the necessity of tedious

computation. However, the curious reader may find a rigorous demonstration of such computational techniques in [14].

Existing research assures us of the overall statistical significance of depression status and age as predictors of RARs. Accordingly, we are most interested in examining the timing of statistically significant depression, age group, and interaction effects as well as the overall statistical significance of the interaction. We can accomplish the former in a familiar manner by analyzing a set of estimated coefficient functions and their pointwise confidence intervals. However, to determine the overall significance of the interaction between depression and age group, we must turn to permutation F-testing. There are four main steps to performing the test. First, we assume a null model. Here, we use the main effects model. Second, we permute in such a manner that preserves the null. This yields the empirical distribution. Third, for a significance level of .05, we choose a critical value such that in five percent of cases we reject the null. In our case, this is to say that we choose a critical value such that there are five of one hundred F-statistic curves for which the maximum value exceeds the critical value chosen. Last, we check to see if the original curve is among those five F-statistic curves. If this is the case, we reject the null in favor of our full model.

4.0 Results

4.1 Exploratory Results

As in section 3, we distinguish between two sections of our analysis: the exploratory analysis and the fANOVA. We begin with the exploratory analysis and stipulate that these preliminary results are simply meant to motivate the upcoming, principal section and to perhaps identify some questions that will warrant further study. In this section we present several figures that aim to generally describe the RARs and to highlight certain RAR characteristics common to and different between depressed and nondepressed groups.

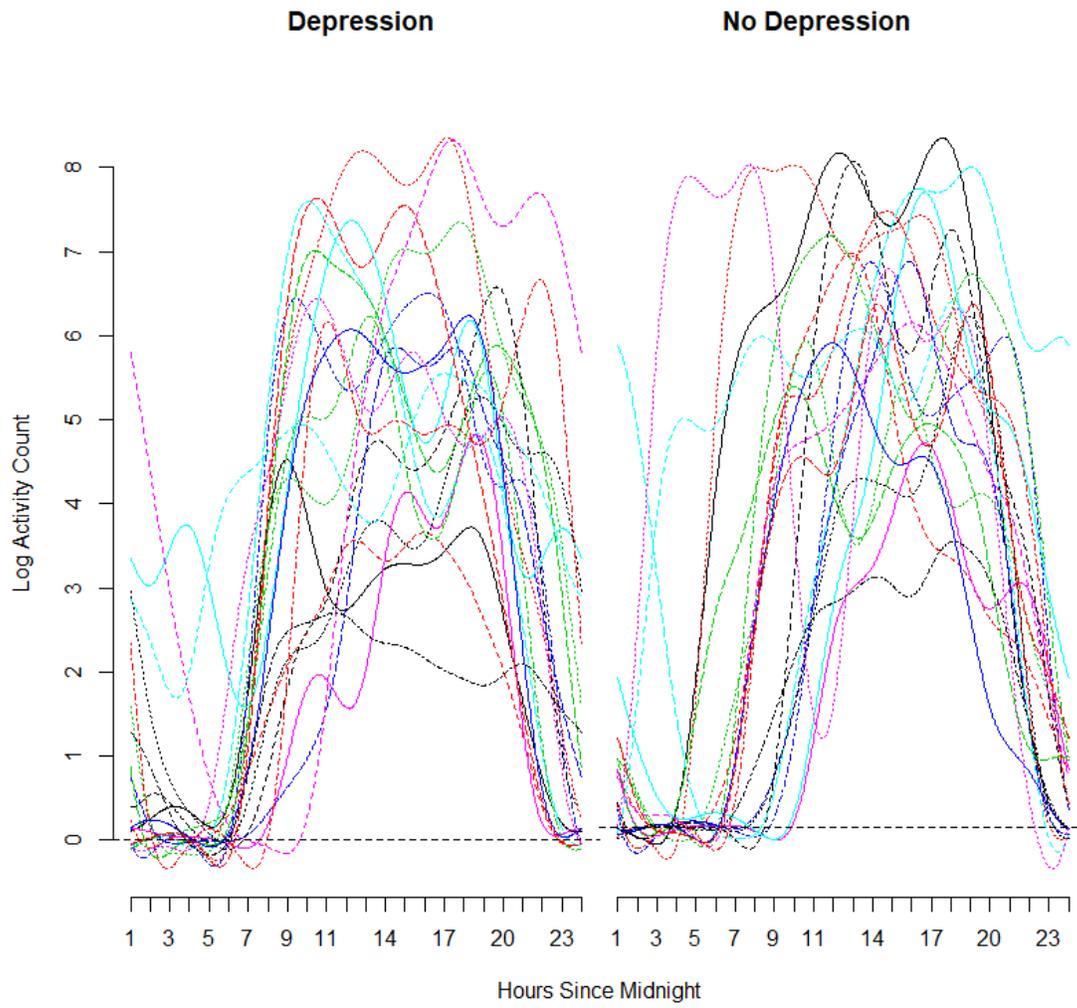


Figure 3 Ten smoothed mean log activity count curves from data subsets by depression status.

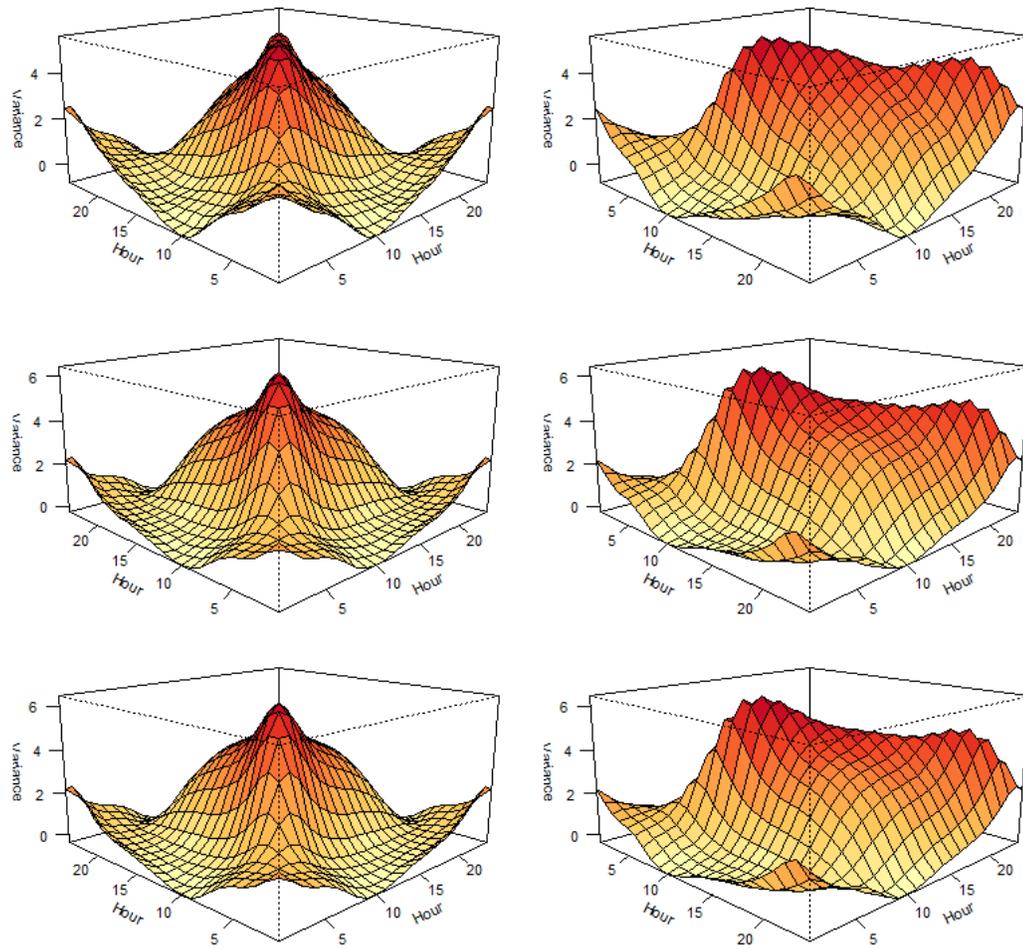


Figure 4 Estimated autocovariance surfaces for the mean log activity count data.

Plots represent the depression subset (top), the no depression subset (middle), and the full data (bottom).

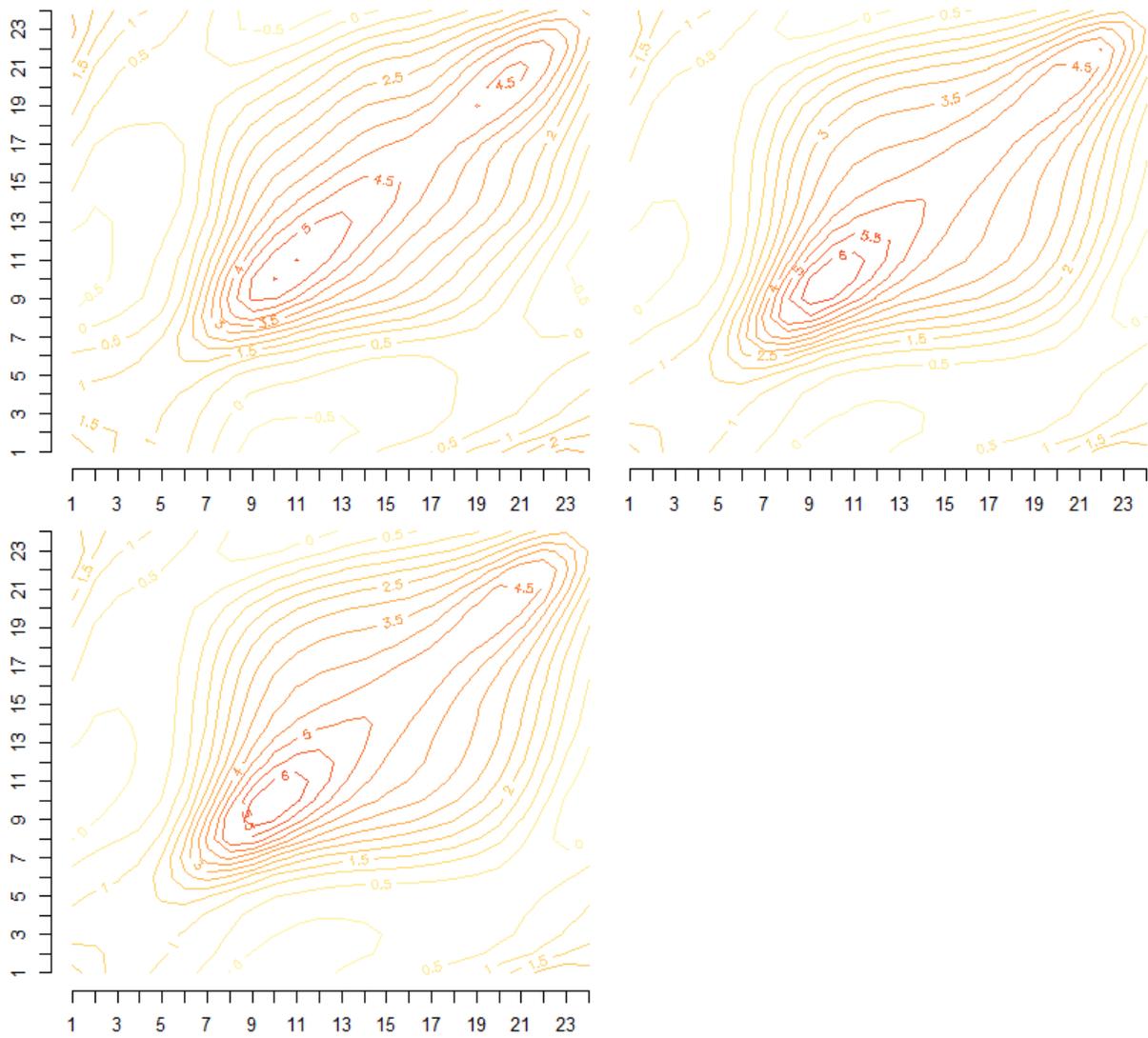


Figure 5 Contour plots of the bivariate variance-covariance surfaces by hours since midnight.

The plots represent the depression subset (top left), the no depression subset (top right), and the full data (bottom).

A full picture of the variation in log activity counts across time is given by Figures 4 and 5, which complement each other in that autocovariation features can easily be identified via perspective plot and traced to a precise time via contour plot. It is no surprise that the plots generated by the no depression subset and the full dataset are almost identical because subjects without depression hold a clear majority in the full dataset. However, we can identify a few incongruities between plots of the depression and no depression subsets. For example, the variances and covariances are mostly greater for the no depression group. Another obvious feature is the variance ridge along the diagonal. We notice that for the no depression group this ridge begins to form earlier than for the depression group and that the covariances descend less steeply from the ridge. The sharp, early morning increase in variance can likely be attributed to the fact that during this period many subjects are still in their rest phase while others have begun their possibly intense morning routines. The difference in steepness of descent from the variance ridge between the two surfaces is minor, and we do not undertake any testing to resolve the issue. However, perhaps some speculation is warranted. If it is the case that higher covariances persist in the no depression group for values farther from the ridge, and if it is not simply the consequence of higher mean log activity counts at those times, one could reasonably conceive that subjects without depression engage in commitments during waking hours that require some similar level of activity for all times during that period. Also, the depression group features two distinct peaks as opposed to one. However, we have reason to be suspicious of late night findings as will be discussed in the next section.

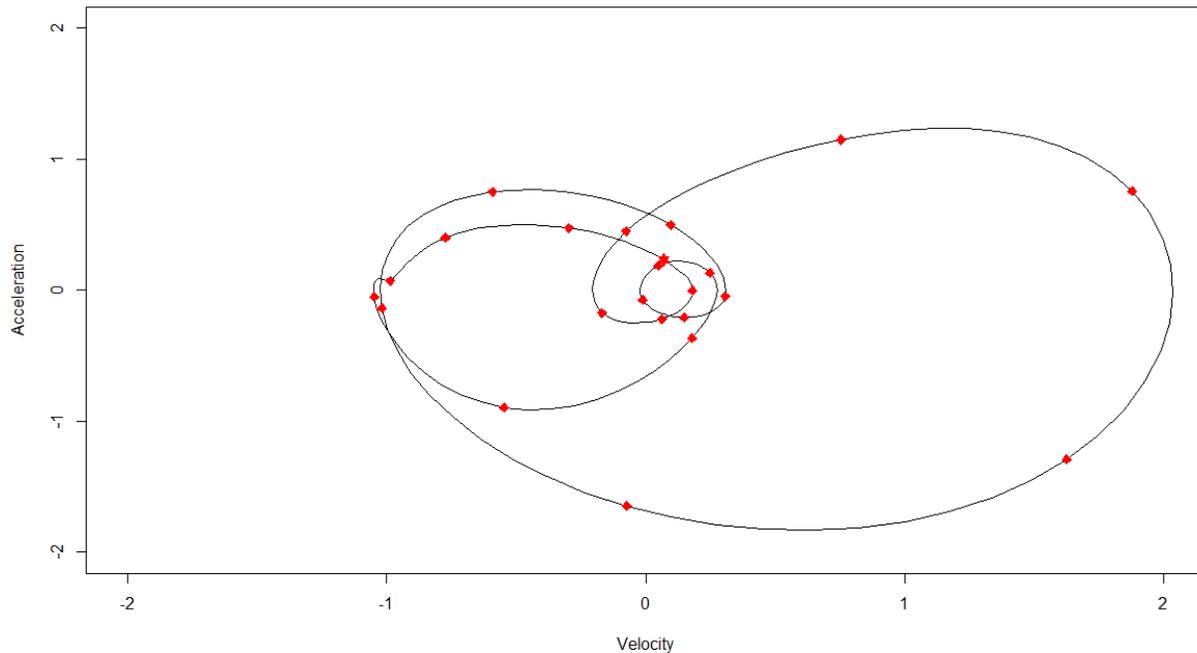


Figure 6 A phase-plane plot for one subject.

The initial and final hours are indicated by the upward and downward arrows. All other hours are indicated by diamonds.

As mentioned in the previous section we can use the phase-plane plot and the concepts of potential and kinetic energy to explain energy transfer within an individual throughout a day. We examine the system's energy through four oscillations. The cycle begins near the center and proceeds clockwise, hovering around the origin as the subject sleeps. Slightly after 5:00 am, the subject increases activity at an increasing rate. This may likely reflect waking and engaging in a morning routine. Slightly after 7:00 am, the system reaches its maximum velocity and activity begins to peak near 9:00 am. Activity falls at a decreasing then increasing rate as the system begins to lose energy and hovers around the center, eventually entering into a cycle from approximately 12:00 am to 6:00 pm, possibly representing sedentary work. Until around 9:00 pm we see the

lowest levels of activity, at which point activity starts to taper and the subject settles before the system again loses energy, and the subject sleeps.

4.2 Functional Analysis of Variance Results

In this section, we present the estimated functional effects of depression status, age, ethnicity, and gender on our functional outcome, mean log activity count, as well as the results of the permuted F-test.

4.2.1 Estimated Coefficient Functions

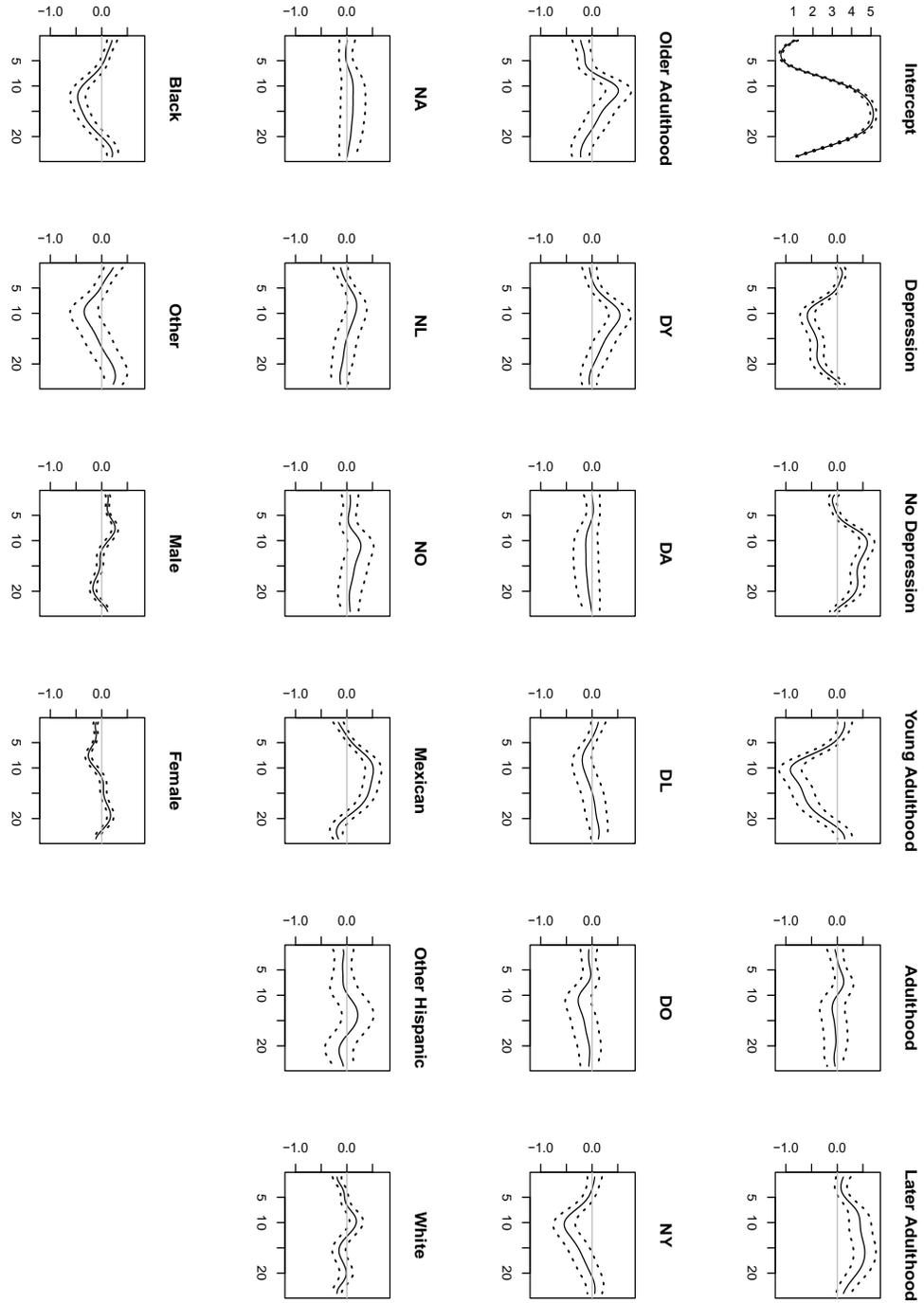


Figure 7 Estimated regression coefficients for predicting log activity count.

The estimated functional covariate effects or regression coefficients represent building blocks that can be additively combined to create a predicted log activity profile for an individual matching a description given by a set of the covariates. It is in this way that we can interpret each of the effects. Furthermore, the dotted lines around the solid line in each frame represent pointwise 95% confidence intervals. Thus, we can determine the significance of each of the effects at a given time. We begin with the depression effect. This effect is significant during all normal waking hours and indicates a considerable depression in the predicted mean log activity profile. The results also show a statistically significant increase in activity is also present in the late night. However, late night findings should be ignored because due to missingness of late night data. As mentioned earlier, subjects were instructed to remove the accelerometer before sleeping. Young adulthood shows a lesser activity during waking hours, while later adulthood shows the opposite. Like the effect for later adulthood, older adulthood indicates an increase in log activity, though no sustained, significant effect is present. The interaction terms do not show any significant effects except in the young adults. This interaction negates much of the dulling effect of the depression term during the late morning and early afternoon. We also observe significant differences among ethnicities and between genders, but we neglect to discuss these in further detail, as both covariates were included simply for the sake of generalizability.

4.2.2 Permutation F-test Results

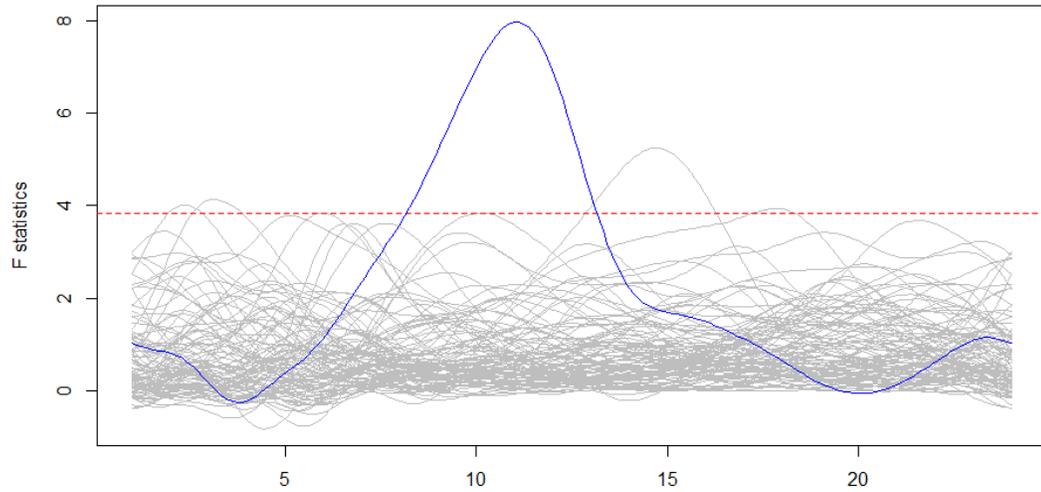


Figure 8 Permutation F-statistics Plot.

From the plots shown above, it is clear that the maximum value of the original F-statistics curve, plotted in blue, is greater than our chosen critical value. Thus, we conclude that at 95% confidence, the full model captures the data significantly better than the null, or main effects model.

5.0 Conclusions

The broad aims of the present study can be categorized as investigative and demonstrative. We investigated the nature of the relationship between depression status as classified via a PHQ-9 score and an accelerometer-based measure of activity and demonstrated fANOVA as a useful tool in analyzing RAR data. Among our primary interests and most important findings, we found depression status to be a statistically significant predictor of log activity count at a 95% confidence level for all waking hours of the day. Also, we found that young adults are on average statistically significantly less active during waking hours, while individuals in late or older adulthood show the opposite. Furthermore, interaction effects between depression status and age were muted for all age groups except young adulthood for which we see an increase in predicted mean log activity count. That is to say that on average for young adults, the effect of depression during waking hours is less pronounced than for other age groups.

Appendix R Code

```
Sys.setenv(TZ="America/New_York")
Sys.getenv("TZ")
memory.limit(50000)
#####
#### Packages
install.packages("refund")
install.packages("plyr")
install.packages("tidyverse")
devtools::install_github("refunders/refund", ref="devel")
install.packages("fda")
#####
#### Libraries
library(refund)
library(plyr)
library(tidyverse)
library(reshape2)
library(fda)
#####
#### Wrangling
devtools::install_github("andrew-leroux/rnhanesdata")
rawdata <- rnhanesdata::PAXINTEN_D
t.rawdata <- t.data.frame(rawdata) # transpose the data
t.rawdata <- t.rawdata[-c(2:5),] # cut the unnecessary bits
meltdata <- reshape2::melt(t.rawdata[,]) # wide to long format
seqn <- meltdata$value[which(meltdata$Var1 == "SEQN")] # pull the ids (seqn)
seqnrow <- which(meltdata$Var1 == "SEQN") # could be useful later
dt <- seq.POSIXt(ISOdatetime(1,1,1,0,0,0,"GMT"),ISOdatetime(1,1,1,23,59,0,"GMT"), by =
"min") # prepare times
dt <- rep(dt,52185)
inputdata <- matrix(data = NA, nrow = length( meltdata$value[-which(meltdata$Var1 ==
"SEQN")] ), ncol = 4)
inputdata <- as.data.frame(inputdata,stringsAsFactors = FALSE) # dont execute
inputdata[,1] <- as.numeric(as.character(meltdata$value[-which(meltdata$Var1 == "SEQN")])) #
add actigraphy values
inputdata[,2] <- format(dt, format = "%H:%M:%S") # add times
inputdata[,3] <-
as.numeric(as.character(reshape2::melt(sapply(1:length(unique(seqn)),function(d){ rep(d,1440*7
)}))[,3])) # add ids (count)
inputdata[,4] <-
as.numeric(as.character(reshape2::melt(sapply(1:length(seqn),function(d){ rep(seqn[d],1440)}))[
,3])) # add id (seqn)
```

```

colnames(inputdata) <- c("act","time","id","seqn")
ccinputdata <- inputdata[-which(inputdata[,3] %in%
unique(inputdata[,3][which(is.na(inputdata[,1])==TRUE)]),]
ccinputdata <- as.data.frame(ccinputdata,stringsAsFactors = FALSE)
colnames(ccinputdata) <- c("act","time","id","seqn")
ccinputdata$tid <- as.numeric(ccinputdata$tid)
bins <- 24
doa24hb <- matrix(NA,nrow = nid,ncol = bins)
a <- NULL

for (i in 1:nid){

  tryCatch({temp <-
ccinputdata[min(which(ccinputdata$tid==i)):max(which(ccinputdata$tid==i)),1:4]
  for (j in 1:bins){

    for (k in 1:7){
      a <- c(a,temp$act[(((1+(1440*(k-1)))+((j-1)*(1440/bins))):(((1440/bins)*j)+(1440*(k-1)))]])
    }

    doa4hb[i,j] <- mean(log(a))
    a <- NULL
  }
},error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
}
doa4hb <- as_tibble(doa4hb)
doa4hb$SEQN <- unique(ccinputdata$seqn)
local_dem <- read.csv("C:/Users/chc289/Documents/demos.csv")
add_dem <- read.csv("C:/Users/chc289/Documents/addtodemos.csv")
dem <- merge(local_dem,add_dem,by="SEQN")
dem <- as_tibble(dem)
dem <- dem[,-c(2,3,4,5,8)]
doa4hb <- doa4hb[which(doa4hb$SEQN %in% dem$SEQN),]
doa4hb$SEQN <- NULL
doa4hb <- as.matrix(doa4hb)
doa4hb <- t(doa4hb)
act_data <- melt(doa4hb)
#####
### FDA
fbasis_30 <- create.fourier.basis(rangeval=c(1, 24), nbasis = 31)
fbasis_30_mat <- eval.basis(evalarg = 1:24, basisobj = fbasis_30)
# we should check which lambda to use and create fdPar
Lcoef <- c(0,(2*pi/24)^2,0)
harmaccelLfd <- vec2Lfd(Lcoef,c(1,24))
loglam <- (seq(-10,10,0.25))
nlam <- length(loglam)

```

```

dfsave <- rep(NA,nlam)
gcvsave <- rep(NA,nlam)
for (ilam in 1:nlam){
  cat(paste('log10 lambda =', loglam[ilam], '\n'))
  lambda <- 10^loglam[ilam]
  fdParobj <- fdPar(fbasis_30,harmaccelLfd,lambda)
  smoothlist <- smooth.basis(1:24,act_data,fdParobj)
  dfsave[ilam] <- smoothlist$df
  gcvsave[ilam] <- sum(as.vector(smoothlist$gcv),na.rm=TRUE)
}
plot(loglam,gcvsave)
lambda <- 1e-1
# done
fdParobj <- fdPar(fbasis_30, harmaccelLfd, lambda)
act.fit <- smooth.basis(1:24,act_data,fdParobj)
act.fd <- act.fit$fd
plot(act.fd)
plotfit.fd(act_data,1:24,act.fd)

actmat <- eval.fd(1:24,act.fd)
actres <- act_data - actmat
actvar1 <- apply(actres^2,1,sum)/4390
actvar2 <- apply(actres^2,2,sum)/24
actvar1fit <- smooth.basis(1:24,actvar1, fdParobj)$fd
plot(actvar1)
lines(actvar1fit) # will get to work sometime
plot(actvar2) # label these by category somehow
# some descriptive stats
meanact_data <- mean(act.fd)
stdfd <- std.fd(act.fd)
# bivariate cov function
actvar.bifd <- var.fd(act.fd)
actvar_mat <- eval.bifd(1:24,1:24,actvar.bifd)
persp(1:24,1:24,actvar_mat,theta=-
45,phi=15,r=3,expand=.5,ticktype='detailed',xlab="Hour",ylab="Hour",zlab="Variance")
contour(1:24,1:24,actvar_mat)
# make phase-plane plot
phaseplanePlot(1:24,act.fd)
deriv.act.fd <- eval.fd(seq(1,24,.1),act.fd,1)
deriv2.act.fd <- eval.fd(seq(1,24,.1),act.fd,2)
plot(deriv.act.fd,deriv2.act.fd,type = "l",xlim=c(-4,4),ylim=c(-4,4))
# setup vectors
a <- which(dem$seqn %in% unique(ccinputdata$seqn))
depressed <- dem$dep[a]
nondepressed <- mod((depressed+1),2)
age1 <- dem$age_grp[a]

```

```

b <- which(age1!=1)
age1[b] <- 0
age2 <- dem$age_grp[a]-1
b <- which(age2!=1)
age2[b] <- 0
age3 <- dem$age_grp[a]-2
b <- which(age3!=1)
age3[b] <- 0
age4 <- dem$age_grp[a]-3
b <- which(age4!=1)
age4[b] <- 0

depage1 <- floor((age1+depressed)/2)
depage2 <- floor((age2+depressed)/2)
depage3 <- floor((age3+depressed)/2)
depage4 <- floor((age4+depressed)/2)
nondepage1 <- floor((age1+nondepressed)/2)
nondepage2 <- floor((age2+nondepressed)/2)
nondepage3 <- floor((age3+nondepressed)/2)
nondepage4 <- floor((age4+nondepressed)/2)

```

```

eth1 <- dem$RIDRETH1[a]
b <- which(eth1!=1)
eth1[b] <- 0
eth2 <- dem$RIDRETH1[a]-1
b <- which(eth2!=1)
eth2[b] <- 0
eth3 <- dem$RIDRETH1[a]-2
b <- which(eth3!=1)
eth3[b] <- 0
eth4 <- dem$RIDRETH1[a]-3
b <- which(eth4!=1)
eth4[b] <- 0
eth5 <- dem$RIDRETH1[a]-4
b <- which(eth5!=1)
eth5[b] <- 0

```

```

gen1 <- dem$RIAGENDR.y[a]
b <- which(gen1!=1)
gen1[b] <- 0
gen2 <- dem$RIAGENDR.y[a]-1
b <- which(gen2!=1)
gen2[b] <- 0

```

```

#

depstat <- c("Intercept", "Depression", "No Depression", "Young Adulthood", "Adulthood", "Later
Adulthood",
            "Older Adulthood", "DY", "DA", "DL", "DO", "NY", "NA", "NL", "NO", "Mexican", "Other
Hispanic", "White", "Black", "Other", "Male", "Female")
p <- 22
deplist <- vector("list", p)
deplist[[1]] <- c(rep(1,4390),0,0,0,0,0)
deplist[[2]] <- c(depressed,1,0,0,0,0)
deplist[[3]] <- c(nondepressed,1,0,0,0,0)
deplist[[4]] <- c(age1,0,1,0,0,0)
deplist[[5]] <- c(age2,0,1,0,0,0)
deplist[[6]] <- c(age3,0,1,0,0,0)
deplist[[7]] <- c(age4,0,1,0,0,0)
deplist[[8]] <- c(depage1,0,0,1,0,0)
deplist[[9]] <- c(depage2,0,0,1,0,0)
deplist[[10]] <- c(depage3,0,0,1,0,0)
deplist[[11]] <- c(depage4,0,0,1,0,0)
deplist[[12]] <- c(nondepage1,0,0,1,0,0)
deplist[[13]] <- c(nondepage2,0,0,1,0,0)
deplist[[14]] <- c(nondepage3,0,0,1,0,0)
deplist[[15]] <- c(nondepage4,0,0,1,0,0)
deplist[[16]] <- c(eth1,0,0,0,1,0)
deplist[[17]] <- c(eth2,0,0,0,1,0)
deplist[[18]] <- c(eth3,0,0,0,1,0)
deplist[[19]] <- c(eth4,0,0,0,1,0)
deplist[[20]] <- c(eth5,0,0,0,1,0)
deplist[[21]] <- c(gen1,0,0,0,0,1)
deplist[[22]] <- c(gen2,0,0,0,0,1)
#

modmat <-
cbind(deplist[[1]][1:4390],deplist[[2]][1:4390],deplist[[3]][1:4390],deplist[[4]][1:4390],deplist[[
5]][1:4390],deplist[[6]][1:4390],deplist[[7]][1:4390],

deplist[[8]][1:4390],deplist[[9]][1:4390],deplist[[10]][1:4390],deplist[[11]][1:4390],deplist[[12]]
[1:4390],deplist[[13]][1:4390],deplist[[14]][1:4390],

deplist[[15]][1:4390],deplist[[16]][1:4390],deplist[[17]][1:4390],deplist[[18]][1:4390],deplist[[1
9]][1:4390],deplist[[20]][1:4390],deplist[[21]][1:4390], deplist[[22]][1:4390])
constraints_check <- matrix(c(0,0,1,0,1,0,0,1,0,1,0,1,0,1),nrow=2,ncol=7)

```



```
mymat0[4,] <- c(rep(0,12),1,1)
```

Bibliography

- [1] Ancoli-Israel, S., Klauber, M. R., Jones, D. W., Kripke, D. F., Martin, J., Mason, W., Pat-Horenczyk, R., & Fell, R. (1997). Variations in circadian rhythms of activity, sleep, and light exposure related to dementia in nursing-home patients. *Sleep*, 20(1), 18-23.
- [2] Bůžková, P., Lumley, T., & Rice, K. (2011). Permutation and parametric bootstrap tests for gene-gene and gene-environment interactions. *Annals of human genetics*, 75(1), 36–45. <https://doi.org/10.1111/j.1469-1809.2010.00572.x>
- [3] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire). Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, [2005-2006].
- [4] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, [2005-2006].
- [5] Craven, P., Wahba, G. (1978). Smoothing noisy data with spline functions. *Numer. Math.* 31, 377–400. <https://doi.org/10.1007/BF01404567>
- [6] Goldsmith, J., Zipunnikov, V., & Schrack, J. (2015). Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2), 344–353. <https://doi.org/10.1111/biom.12278>
- [7] Harper, D. G., Stopa, E. G., McKee, A. C., Satlin, A., Harlan, P., Goldstein, R., & Volicer, L. (2001). Differential circadian rhythm disturbances in men with Alzheimer disease and frontotemporal degeneration. *Archives of General Psychiatry*, 58(4), 353–360. <https://doi.org/10.1001/archpsyc.58.4.353>
- [8] Hatfield, Catherine & Herbert, Joe & Van Someren, Eus J W & Hastings, M.H. (2004). Disrupted daily activity/rest cycles in relation to daily cortisol rhythms of home-dwelling patients with early Alzheimer's dementia. *Brain: a journal of neurology*. 127. 1061-74. <https://doi.org/10.1093/brain/awh129>.
- [9] Hori, H., Koga, N., Hidese, S., Nagashima, A., Kim, Y., Higuchi, T., & Kunugi, H. (2016). 24-h activity rhythm and sleep in depressed outpatients. *Journal of psychiatric research*, 77, 27–34. <https://doi.org/10.1016/j.jpsychires.2016.02.022>

- [10] Krafty, R.T. (2015, October). *Rest-activity rhythms and future depression: Prospective association and biological speculation*. Presentation as an installment in the University of Pittsburgh GSPH Grand Rounds series, Pittsburgh, PA.
- [11] Luik, A. I., Zuurbier, L. A., Hofman, A., Van Someren, E. J., Ikram, M. A., & Tiemeier, H. (2015). Associations of the 24-h activity rhythm and sleep with cognition: a population-based study of middle-aged and elderly persons. *Sleep medicine*, *16*(7), 850–855. <https://doi.org/10.1016/j.sleep.2015.03.012>
- [12] Luik, A. I., Zuurbier, L. A., Hofman, A., Van Someren, E. J., & Tiemeier, H. (2013). Stability and fragmentation of the activity rhythm across the sleep-wake cycle: the importance of age, lifestyle, and mental health. *Chronobiology international*, *30*(10), 1223–1230. <https://doi.org/10.3109/07420528.2013.813528>
- [13] Maglione, J. E., Ancoli-Israel, S., Peters, K. W., Paudel, M. L., Yaffe, K., Ensrud, K. E., Tranah, G. J., Stone, K. L., & Study of Osteoporotic Fractures Research Group (2014). Depressive symptoms and circadian activity rhythm disturbances in community-dwelling older women. *The American journal of geriatric psychiatry: official journal of the American Association for Geriatric Psychiatry*, *22*(4), 349–361. <https://doi.org/10.1016/j.jagp.2012.09.003>
- [14] Ramsay, J., & Silverman, B. (2006). *Functional Data Analysis* (Springer Series in Statistics). New York, NY: Springer Science+Business Media, LLC.
- [15] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- [16] Robillard, R., Hermens, D. F., Naismith, S. L., White, D., Rogers, N. L., Ip, T. K., Mullin, S. J., Alvares, G. A., Guastella, A. J., Smith, K. L., Rong, Y., Whitwell, B., Southan, J., Glozier, N., Scott, E. M., & Hickie, I. B. (2015). Ambulatory sleep-wake patterns and variability in young people with emerging mental disorders. *Journal of psychiatry & neuroscience: JPN*, *40*(1), 28–37. <https://doi.org/10.1503/jpn.130247>
- [17] Robillard, R., Naismith, S. L., Smith, K. L., Rogers, N. L., White, D., Terpening, Z., Ip, T. K., Hermens, D. F., Whitwell, B., Scott, E. M., & Hickie, I. B. (2014). Sleep-wake cycle in young and older persons with a lifetime history of mood disorders. *PloS one*, *9*(2), e87763. <https://doi.org/10.1371/journal.pone.0087763>
- [18] Saito, Y., Kume, Y., Kodama, A., Sato, K., & Yasuba, M. (2018). The association between circadian rest-activity patterns and the behavioral and psychological symptoms depending on the cognitive status in Japanese nursing-home residents. *Chronobiology international*, *35*(12), 1670–1679. <https://doi.org/10.1080/07420528.2018.1505752>
- [19] Satlin, A., Volicer, L., Stopa, E. G., & Harper, D. (1995). Circadian locomotor activity and core-body temperature rhythms in Alzheimer's disease. *Neurobiology of Aging*, *16*(5), 765–771. [https://doi.org/10.1016/0197-4580\(95\)00059-N](https://doi.org/10.1016/0197-4580(95)00059-N)

- [20] Smagula, S. F., Gujral, S., Capps, C. S., & Krafty, R. T. (2019). A Systematic Review of Evidence for a Role of Rest-Activity Rhythms in Dementia. *Frontiers in psychiatry*, *10*, 778. <https://doi.org/10.3389/fpsy.2019.00778>
- [21] Smagula S. F. (2016). Opportunities for clinical applications of rest-activity rhythms in detecting and preventing mood disorders. *Current opinion in psychiatry*, *29*(6), 389–396. <https://doi.org/10.1097/YCO.0000000000000283>
- [22] Tranah, G. J., Blackwell, T., Stone, K. L., Ancoli-Israel, S., Paudel, M. L., Ensrud, K. E., Cauley, J. A., Redline, S., Hillier, T. A., Cummings, S. R., Yaffe, K., & SOF Research Group (2011). Circadian activity rhythms and risk of incident dementia and mild cognitive impairment in older women. *Annals of neurology*, *70*(5), 722–732. <https://doi.org/10.1002/ana.22468>
- [23] van Someren, E. J. W., Hagebeuk, E. E. O., Lijzenga, C., Scheltens, P., de Rooij, S. E. J. A., Jonker, C., Pot, A.-M., Mirmiran, M., & Swaab, D. F. (1996). Circadian rest-activity rhythm disturbances in Alzheimer's disease. *Biological Psychiatry*, *40*(4), 259–270. [https://doi.org/10.1016/0006-3223\(95\)00370-3](https://doi.org/10.1016/0006-3223(95)00370-3)
- [24] Witting, W., Kwa, I. H., Eikelenboom, P., Mirmiran, M., & Swaab, D. F. (1990). Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. *Biological Psychiatry*, *27*(6), 563–572. [https://doi.org/10.1016/0006-3223\(90\)90523-5](https://doi.org/10.1016/0006-3223(90)90523-5)
- [25] Xiao, L., Huang, L., Schrack, J. A., Ferrucci, L., Zipunnikov, V., & Crainiceanu, C. M. (2015). Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics (Oxford, England)*, *16*(2), 352–367. <https://doi.org/10.1093/biostatistics/kxu045>