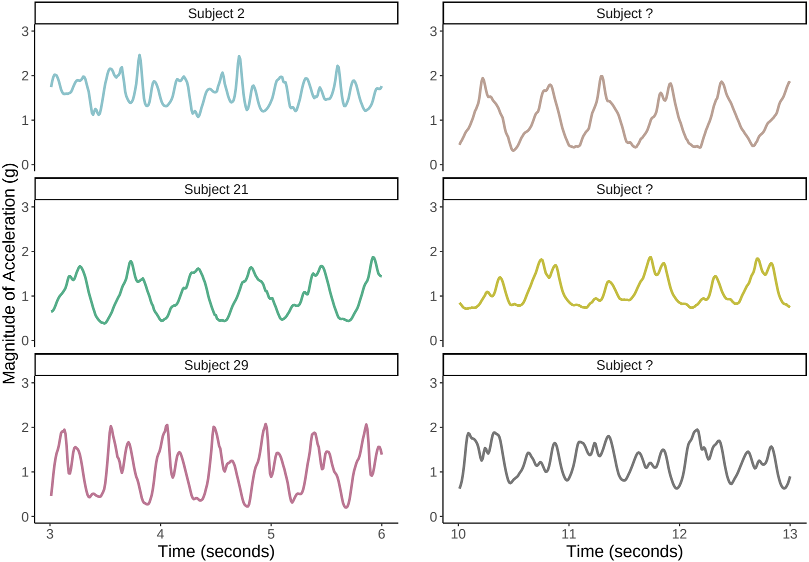
Fingerprinting Walking using High Density Accelerometer Data

ENAR 2024

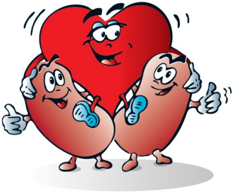
Lily Koffman

The Problem: Can We Identify an Individual From Their Walking?



Why Do We Care?

Why Do We Care?



Approach

- Obtain the empirical joint distribution of acceleration, and lag acceleration for all possible lags (which can be represented as a series of images)

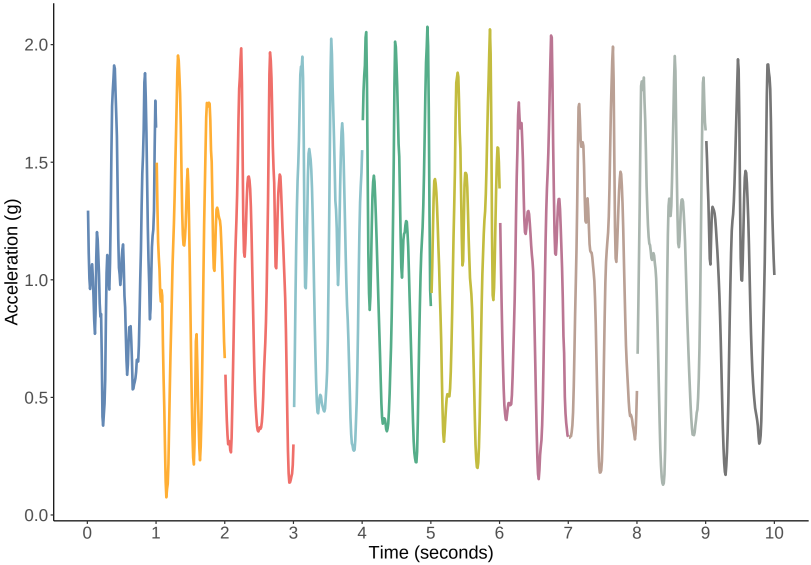
Approach

- Obtain the empirical joint distribution of acceleration, and lag acceleration for all possible lags (which can be represented as a series of images)
    - **Image partitioning**: compute summaries of the joint distribution, use summaries to predict identity

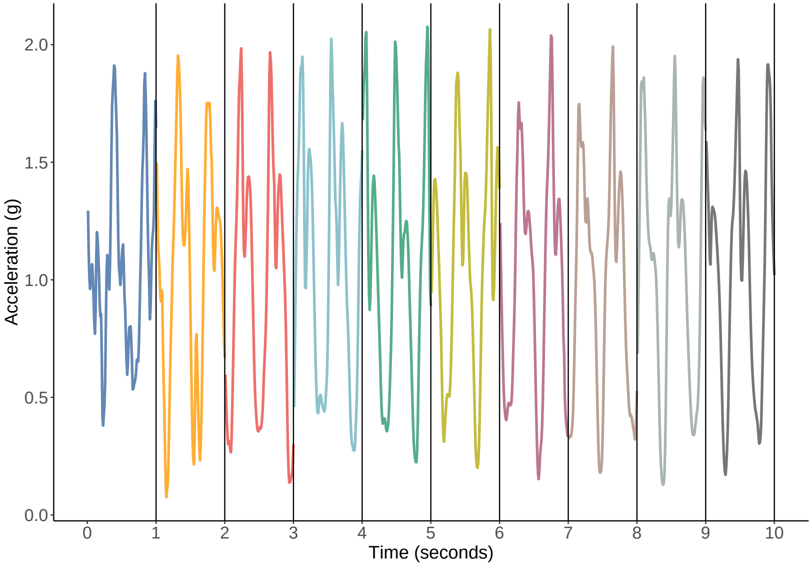
Approach

- Obtain the empirical joint distribution of acceleration, and lag acceleration for all possible lags (which can be represented as a series of images)
  - **Image partitioning**: compute summaries of the joint distribution, use summaries to predict identity
  - **Functional regression**: use joint distribution in trivariate functional regression to predict identity

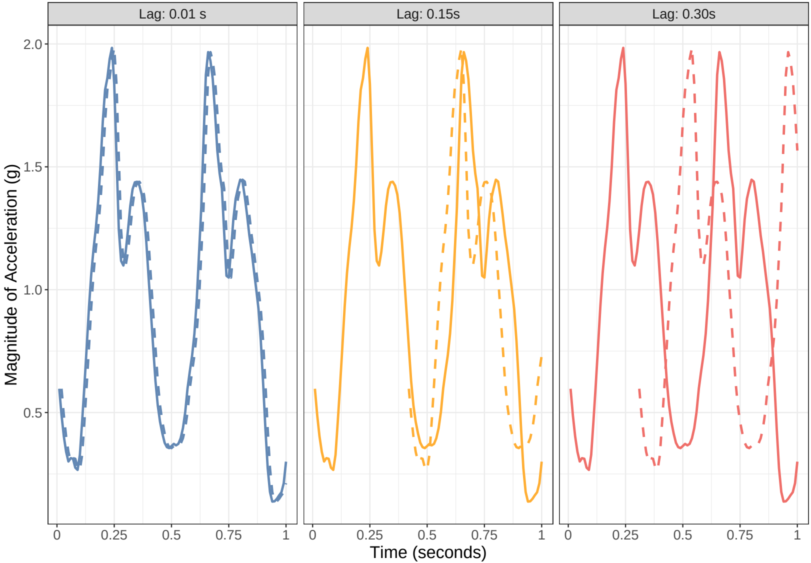
Obtaining the joint distribution: segment the data



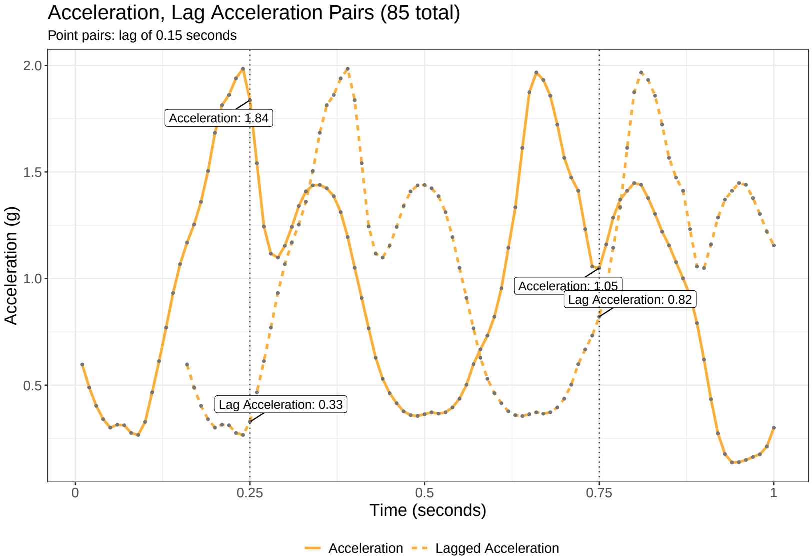
Obtaining the joint distribution: segment the data



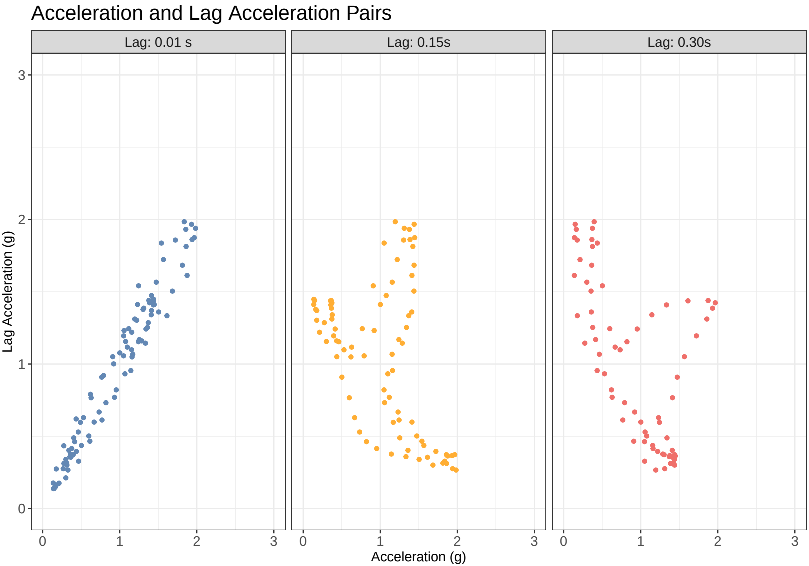
Obtaining the joint distribution: examine acceleration, lag acceleration for each segment and lag



Obtaining the joint distribution: examine acceleration, lag acceleration for each segment and lag



Obtaining the joint distribution: examine acceleration, lag acceleration for each segment and lag



Obtaining the joint distribution: examine acceleration, lag acceleration for each segment and lag

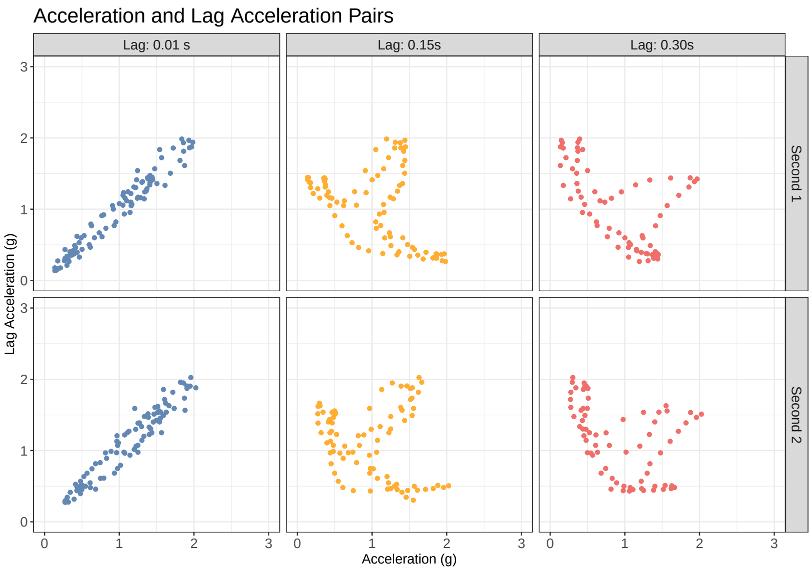


Image partitioning: partition grid into 2D cells

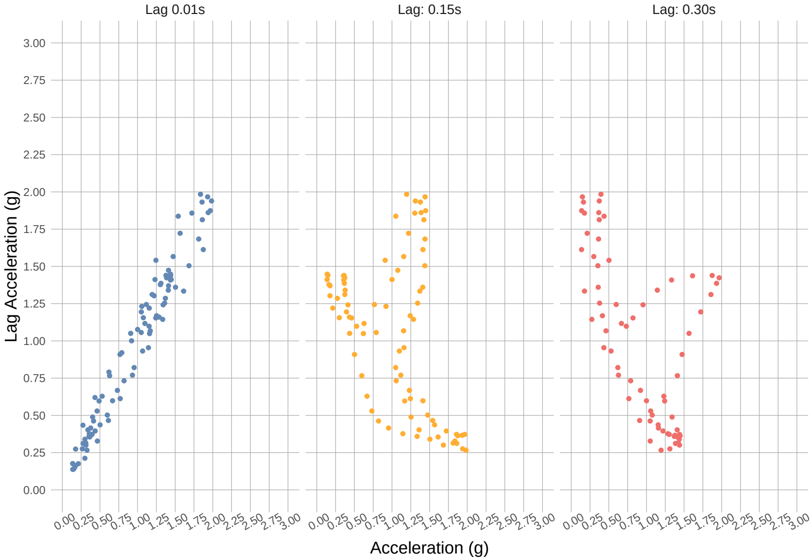


Image partitioning: partition grid into 2D cells: count points in each cell

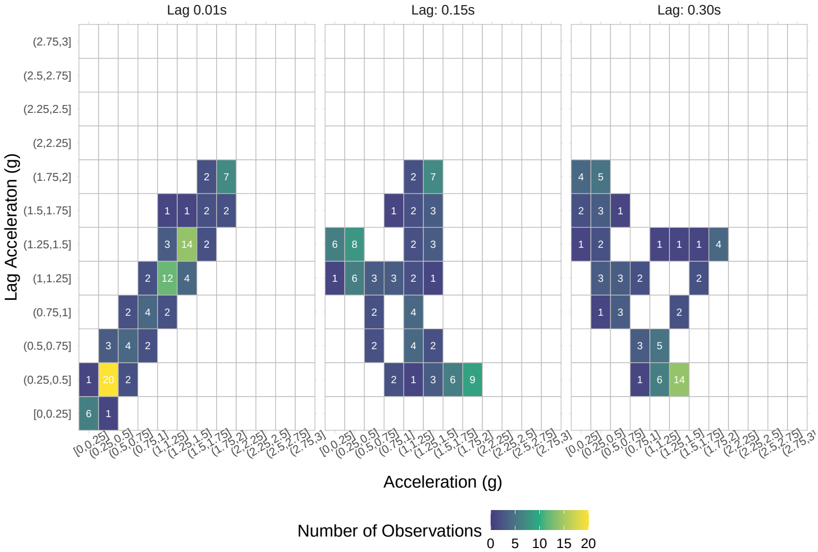


Image partitioning: partition grid into 2D cells: count points in each cell

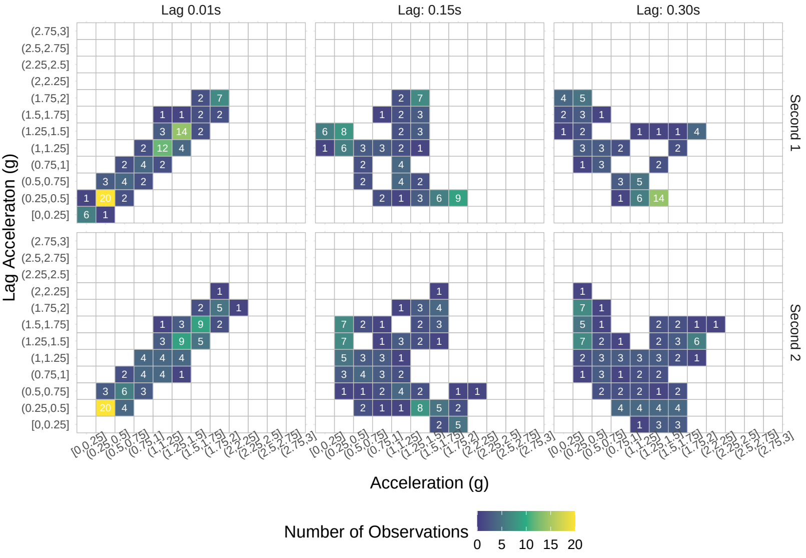


Image partitioning: partition grid into 2D cells: select predictors from grid cells

- We have 144 cells for each lag and 99 possible lags – too many predictors!

Image partitioning: partition grid into 2D cells: select predictors from grid cells

- We have 144 cells for each lag and 99 possible lags – too many predictors!
- Exploratory analyses indicated three lags (0.15, 0.30, 0.45s) performed well

Image partitioning: partition grid into 2D cells: select predictors from grid cells

- We have 144 cells for each lag and 99 possible lags – too many predictors!
- Exploratory analyses indicated three lags (0.15, 0.30, 0.45s) performed well
- $144 * 3$ potential predictors

Image partitioning: partition grid into 2D cells: select predictors from grid cells

- We have 144 cells for each lag and 99 possible lags – too many predictors!
- Exploratory analyses indicated three lags (0.15, 0.30, 0.45s) performed well
- $144 * 3$ potential predictors
- Remove predictors with near zero variance or few unique values

Image partitioning: partition grid into 2D cells: fit models

- We have transformed abstract problem into well-defined classification problem

Image partitioning: partition grid into 2D cells: fit models

- We have transformed abstract problem into well-defined classification problem
- Model:

$$Y_{ij}^{i_0} | X_{ij1}, \dots, X_{ijG}$$

  Where X_{ijg} is number of acceleration, lag acceleration pairs for subject i in grid cell g at second j, $Y_{ij}^{i_0} = 1$ if subject $i = i_0$, and 0 otherwise

Image partitioning: partition grid into 2D cells: fit models

- We have transformed abstract problem into well-defined classification problem
- Model:

$$Y_{ij}^{i_0} | X_{ij1}, \dots, X_{ijG}$$

  Where X_{ijg} is number of acceleration, lag acceleration pairs for subject i in grid cell g at second j, $Y_{ij}^{i_0} = 1$ if subject $i = i_0$, and 0 otherwise
- Use one vs. rest classification (separate model for each individual)

Image partitioning: partition grid into 2D cells: fit models

- We have transformed abstract problem into well-defined classification problem
- Model:

$$Y_{ij}^{i_0} | X_{ij1}, \dots, X_{ijG}$$

  Where X_{ijg} is number of acceleration, lag acceleration pairs for subject i in grid cell g at second j, $Y_{ij}^{i_0} = 1$ if subject $i = i_0$, and 0 otherwise
- Use one vs. rest classification (separate model for each individual)
- Machine learning using `tidymodels`; logistic regression . . .
- For logistic regression models, use correlation and multiplicity adjusted (CMA) confidence intervals for coefficients to identify grid cells that are most predictive of identity

Functional regression

- Instead of summarizing joint distribution, use functional regression of the form:

$$\text{logit}\{p_{ij}^{i_0}\} = \int_{s,u} F\{v_{ij}(s-u), v_{ij}(s), u\} ds du$$

Where $Y_{ij}^{i_0} \sim$ Bernoulli$(p_{ij}^{i_0})$, $u = 1, \ldots, S-1 = 99$, $s = u+1, \ldots, S = 100$, $v_{ij}(s-u)$ is acceleration for subject i, second j, at $s-u$ and $v_{ij}(s)$ is acceleration for subject i, second j, at s.

Functional regression

- Instead of summarizing joint distribution, use functional regression of the form:

$$\text{logit}\{p_{ij}^{i_0}\} = \int_{s,u} F\{v_{ij}(s-u), v_{ij}(s), u\} ds du$$

Where $Y_{ij}^{i_0} \sim \text{Bernoulli}(p_{ij}^{i_0})$, $u = 1, \dots, S-1 = 99$, $s = u+1, \dots, S = 100$, $v_{ij}(s-u)$ is acceleration for subject i, second j, at $s-u$ and $v_{ij}(s)$ is acceleration for subject i, second j, at s.

- $F(\cdot, \cdot, \cdot)$ takes values at every point in domain of 3D images (acceleration, lag acceleration, and lag)

Functional regression

- Instead of summarizing joint distribution, use functional regression of the form:

$$\text{logit}\{p_{ij}^{i_0}\} = \int_{s,u} F\{v_{ij}(s-u), v_{ij}(s), u\} ds du$$

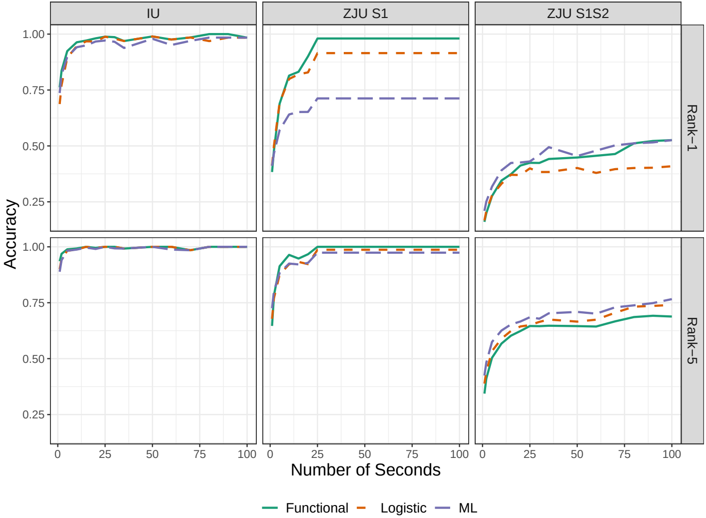
Where $Y_{ij}^{i_0} \sim$ Bernoulli$(p_{ij}^{i_0})$, $u = 1, \dots, S-1 = 99$, $s = u+1, \dots, S = 100$, $v_{ij}(s-u)$ is acceleration for subject i, second j, at $s-u$ and $v_{ij}(s)$ is acceleration for subject i, second j, at s.

- $F(\cdot, \cdot, \cdot)$ takes values at every point in domain of 3D images (acceleration, lag acceleration, and lag)
- Implement model using `mgcv::gam` after manipulating empirical joint distribution into matrices of acceleration, lag acceleration, and lag

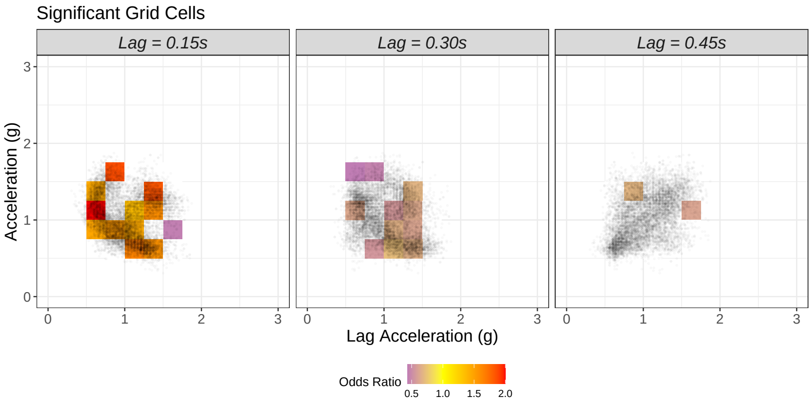
Application

- Two datasets:
  - Indiana Unversity (IU): 32 subjects, 8 min walking per subject
  - Zhejiang University (ZJU): 153 subjects, two trials at least one week and up to six months apart, 1 min walking per subject
    Use for two tasks: within session prediction (train on 75% of seconds in session 1, predict on other 25%)
    Out of session prediction: train on session 1, predict on session 2

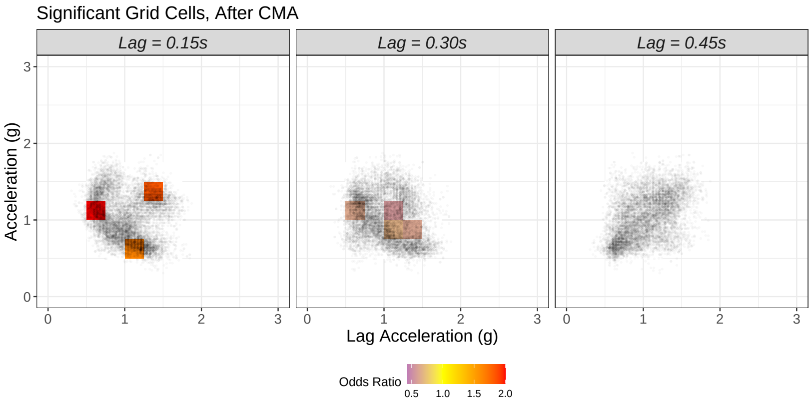
Results: Accuracies over Varying Amount of Testing Data



Inference

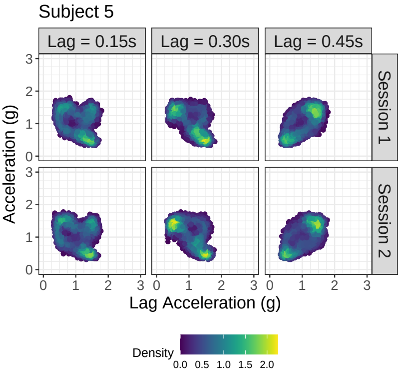

Significant Grid Cells
Lag = 0.15s
Lag = 0.30s
Lag = 0.45s
Acceleration (g)
Lag Acceleration (g)
0
1
2
3
Odds Ratio
0.5
1.0
1.5
2.0

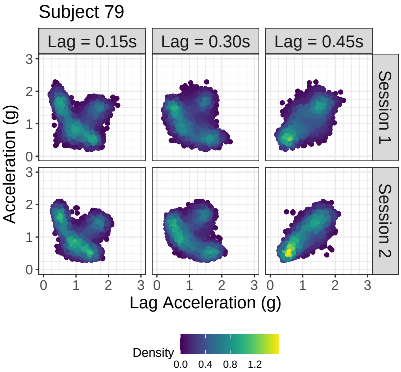
Inference



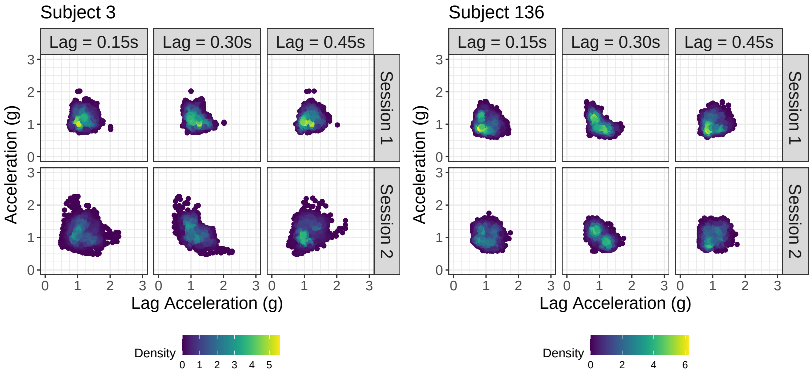
Significant Grid Cells, After CMA
Lag = 0.15s
Lag = 0.30s
Lag = 0.45s
Acceleration (g)
Lag Acceleration (g)
0
1
2
3
Odds Ratio
0.5
1.0
1.5
2.0

Fingerprints: well-predicted subject

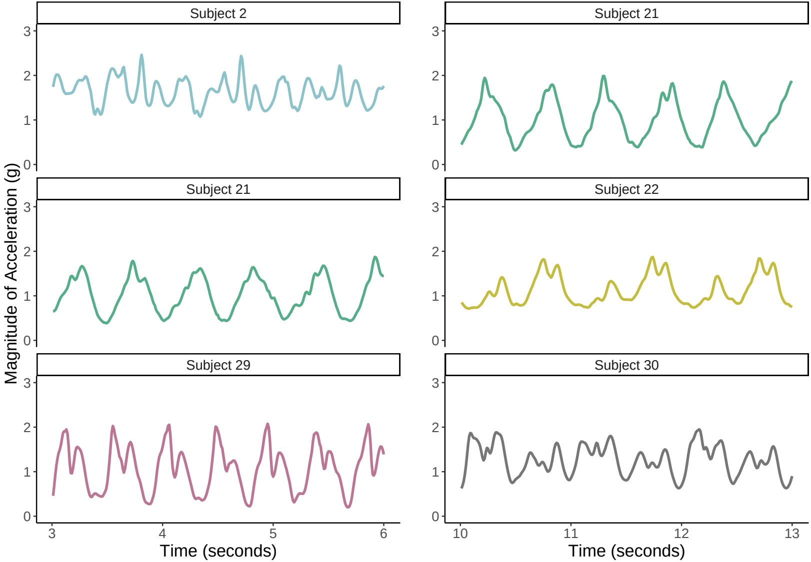




Fingerprints: poorly-predicted subject



Revisiting problem statement



Acknowledgements

- Andrew Leroux, PhD, University of Colorado
- Jaroslaw Harezlak, PhD, Indiana University
- Yan Zhang, ScM, Johns Hopkins Bloomberg School of Public Health
- Ciprian Crainiceanu, PhD, Johns Hopkins Bloomberg School of Public Health