

Segmentation of Accelerometer-derived Time Series

VINCENT VAN HEES, PHD

VINCENTVANHEES@GMAIL.COM

WWW.MOVEMENTDATA.NL

7TH OF APRIL 2020

Introduction



Segmentation helps describe and interpret accelerometer data



Brief overview

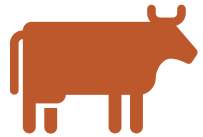


Cut-points and Bouts approach



Hidden Semi-Markov Models

Brief Overview



Raw data:

As used with Deep Neural Networks
and
Template mapping techniques



Signal Features:

e.g. Counts, Acceleration magnitude,
Distribution percentiles



Epochs: Standard length

Defined as: Consecutive- or
overlapping windows



Sojourn/Bouts: Variable length

Defined as: Knowledge-driven rules
and/or Data-driven



Supervised

Given x and y , develop model that can
predict y from x



Unsupervised

Given x , develop model that can
subdivide into data-driven segments
or clusters

Cut-points & Bout/Sojourn approach

Strong need within research community for 'time spent in levels of intensity (energy expenditure)'

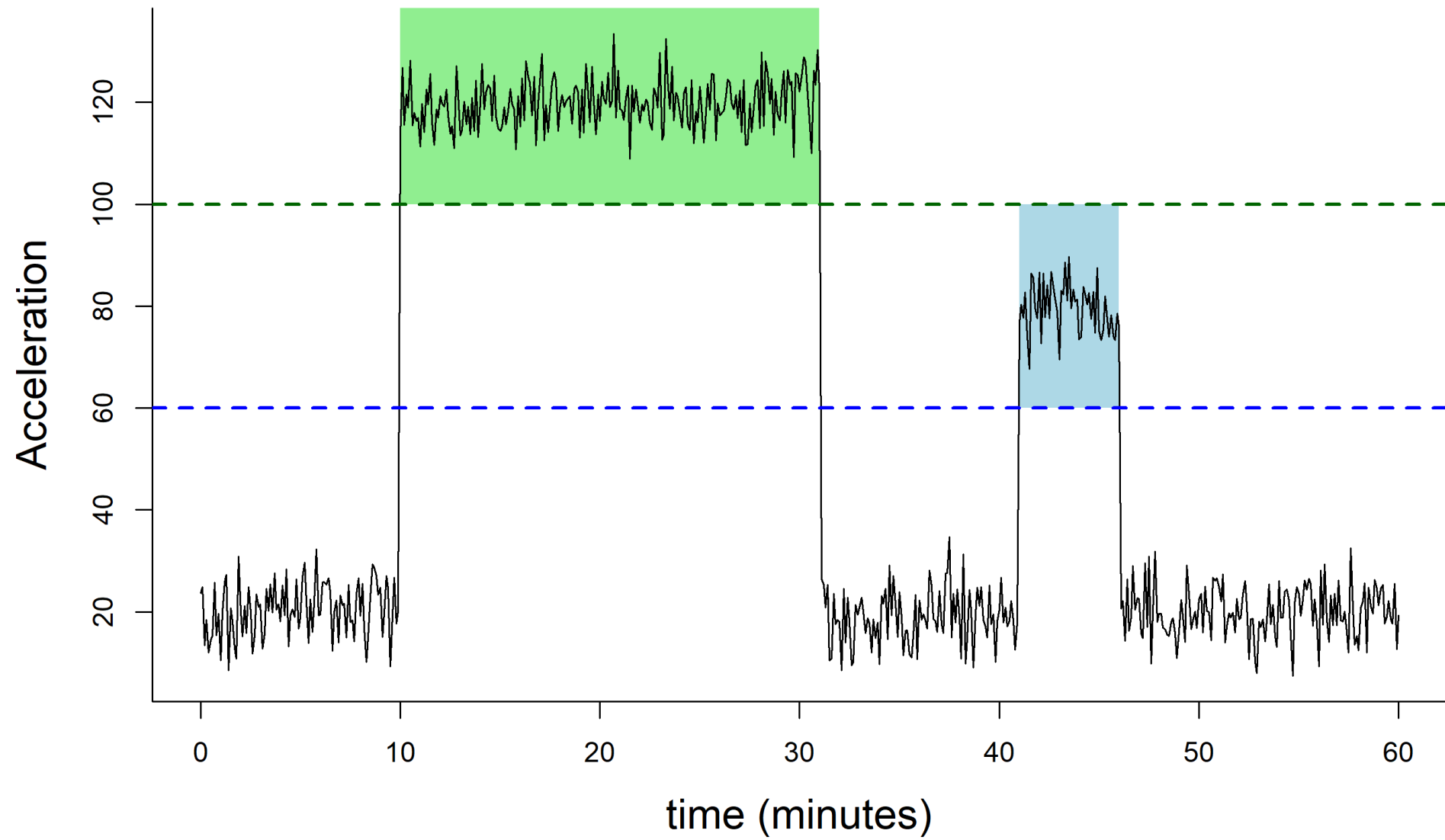
For example: time spent in Moderate or Vigorous Physical Activity (MVPA).

'Self-report method gives it to us, so accelerometer should be able to do better?'

Challenges:

- Acceleration is only one of various determinants of energy expenditure
- Lack of convincing real-life criterion methods
- Lack of clear construct definition

Is it this simple?



Defining what a bout/sojourn is

1. What should the cut-point be?
2. What should the epoch length be?
3. What should minimum duration of bout (sojourn) be?
4. Should we allow for gaps in a bout (sojourn)?
5. Should this be a percentage of the bout duration, an absolute minimum in seconds, or both?
6. Do the first and last epoch need to meet the threshold criteria?
7. Are bout gaps counted towards the time spent in bouts?
8. In what order are the bouts extracted, e.g. what if a 3-minute MVPA bout is part of a 30 minutes inactivity bout?
9. How many bout categories should there be? Inactive, light and MVPA?

Implementation in GGIR

User decides on:

- Acceleration thresholds (3 x)
- Fraction of time for which cut-point criteria need to be met (3 x)
- Bout duration ranges, e.g. [1, 5) [5, 10) and [10, ∞) minutes
- Epoch length

User does NOT decide on:

- Maximum bout gap of 1 minute
- First and last epoch need to meet cut-point criteria
- Number of intensity levels, which are always: inactive, light and MVPA
- Order in which bouts are calculated (1 MVPA; 2 inactive; 3 Light)

Reflections

Useful when interpreted as time spent in acceleration ranges

Unclear whether there is consensus on all these decisions

Unclear how to assess validity

Code validity and documentation probably most important, but never focus in related literature

Other software that segments acc. data

accelerometry: <https://CRAN.R-project.org/package=accelerometry>

acc: <https://CRAN.R-project.org/package=acc>

pawacc: <https://CRAN.R-project.org/package=pawacc>

PASenseWear: <https://CRAN.R-project.org/package=PASenseWear>

Sojourn: <https://CRAN.R-project.org/package=Sojourn>

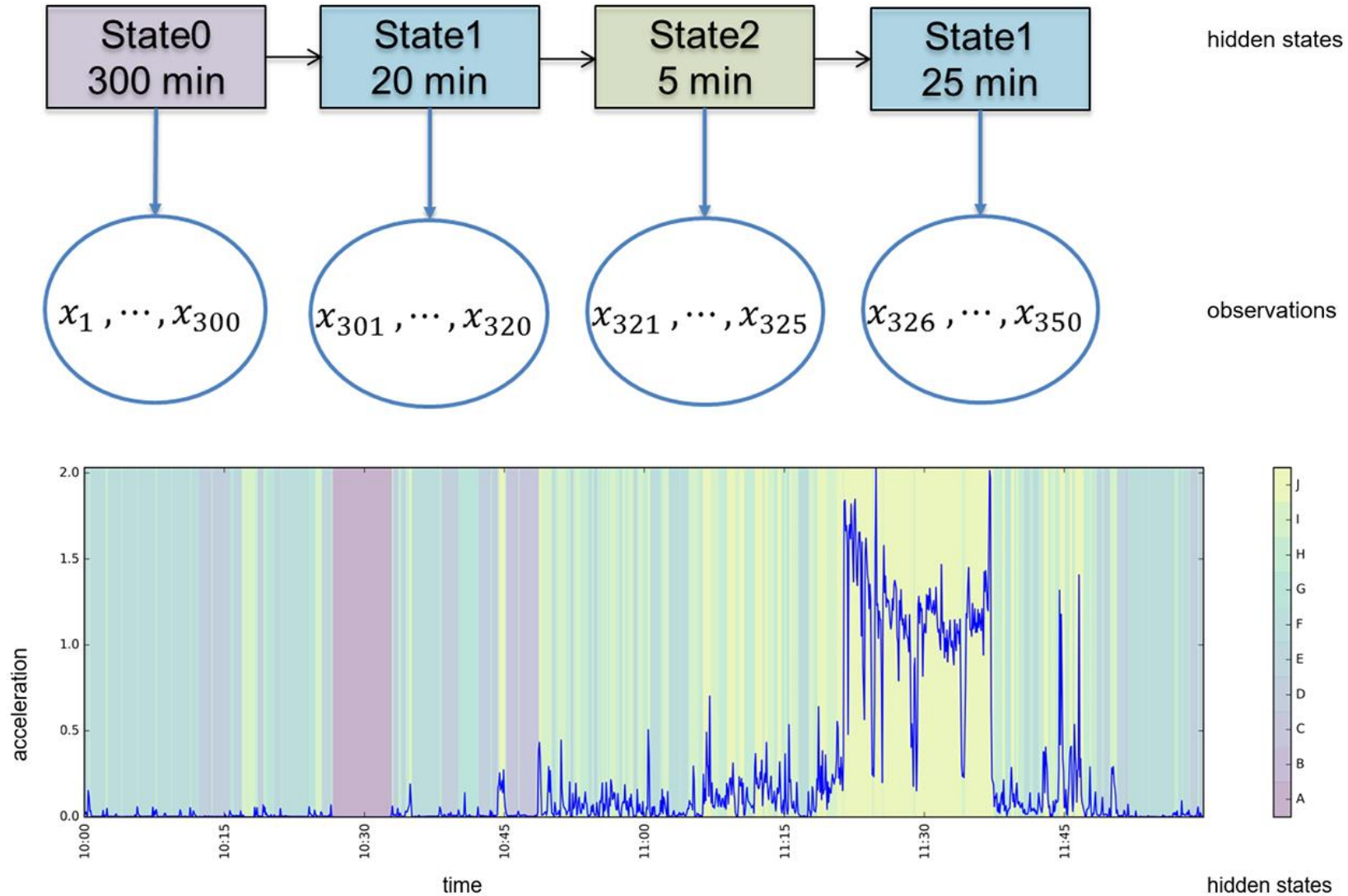
GENEAclassify: <https://CRAN.R-project.org/package=GENEAclassify>

adept: <https://CRAN.R-project.org/package=adept>

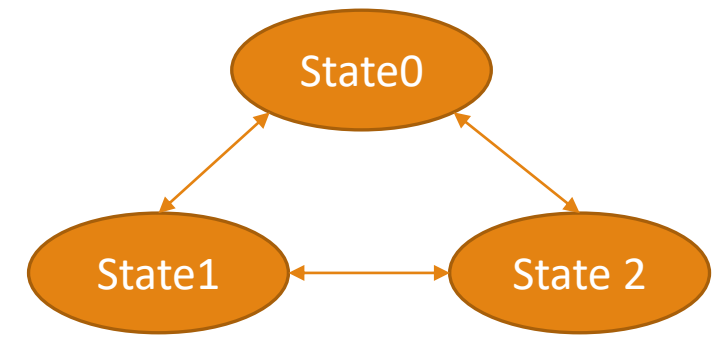
...?

Unsupervised Hidden Semi-Markov Models (HSMM)

Unsupervised learning of abstract clusters (states)



Hidden Semi Markov Model



Initial state i and duration h .

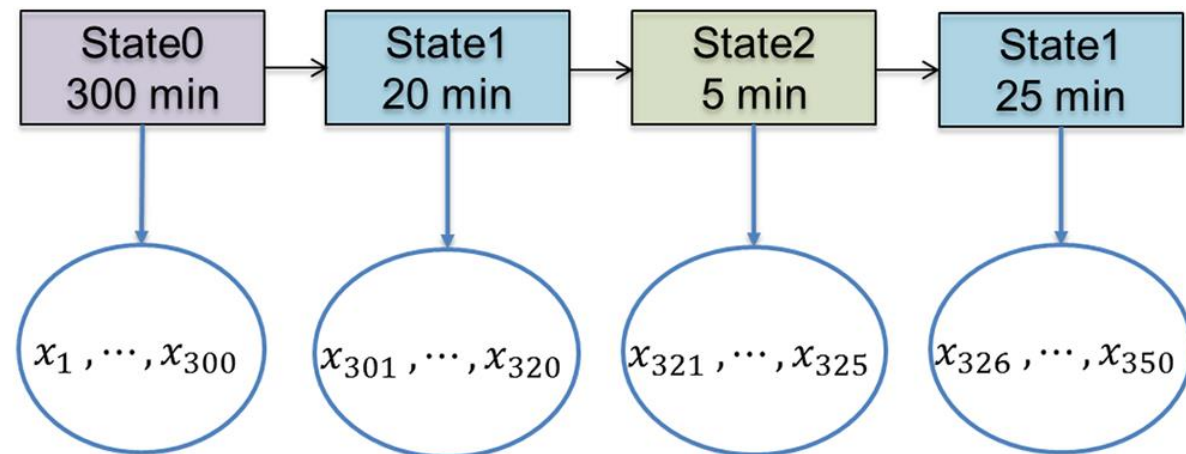
Transition probability matrix $(i, h)(j, d)$ that state i with duration h generates state $j \neq i$ having duration d :

While in state j there will be d observations being **emitted**, independent of time t .

Distribution of durations: We choose Poisson, where each state holds its own lambda parameter.

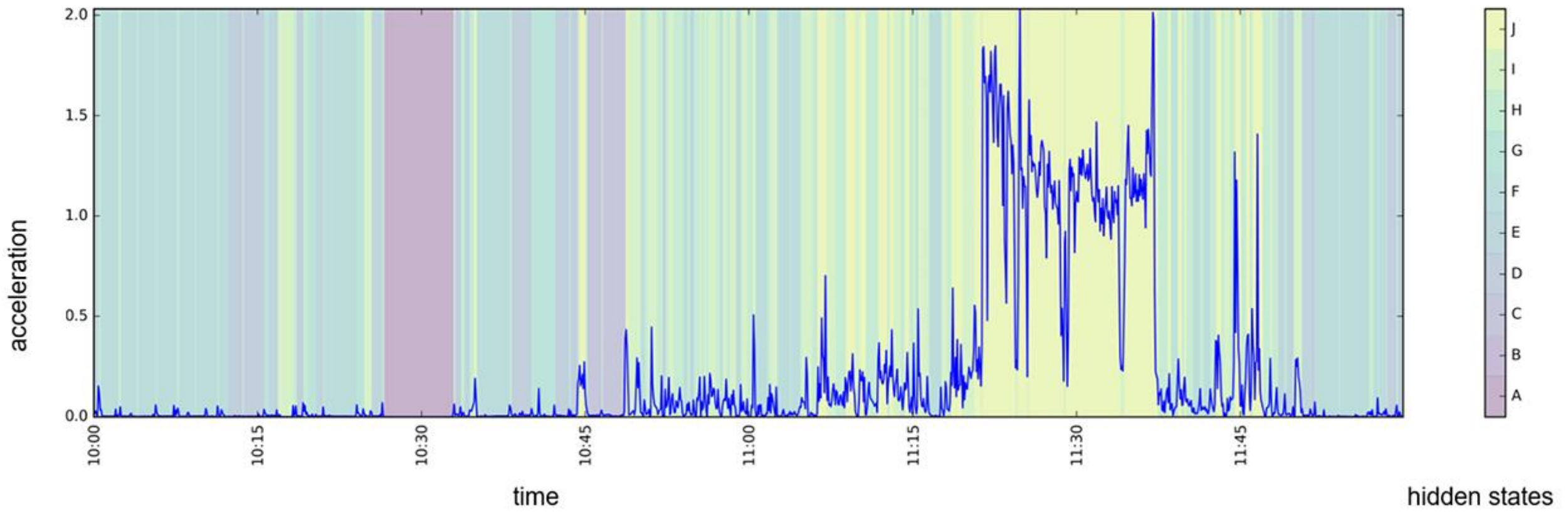
Distribution of observations: We choose multivariate Gaussian distributions.

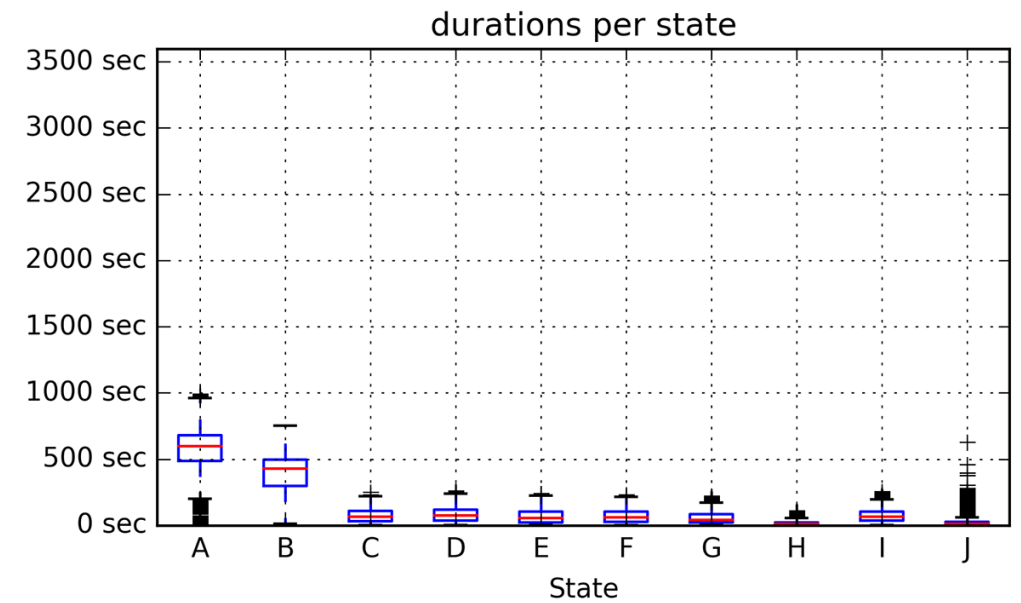
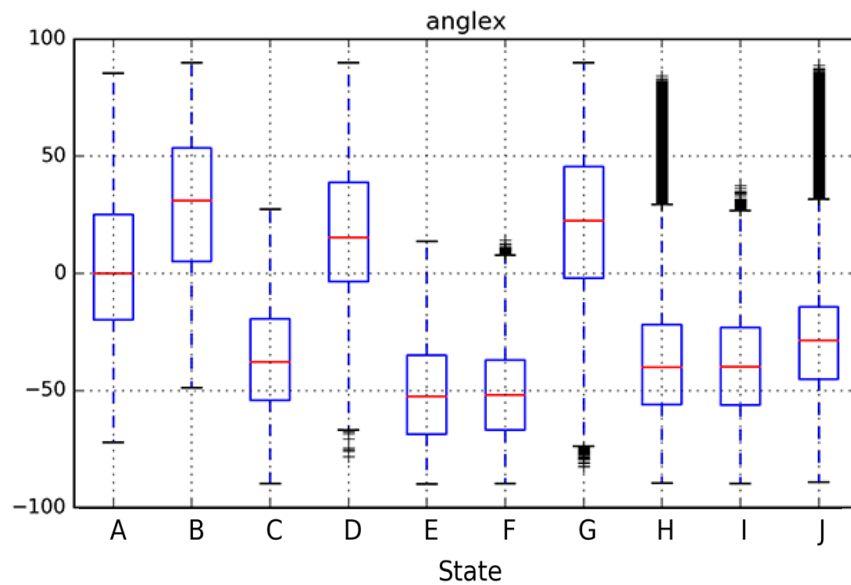
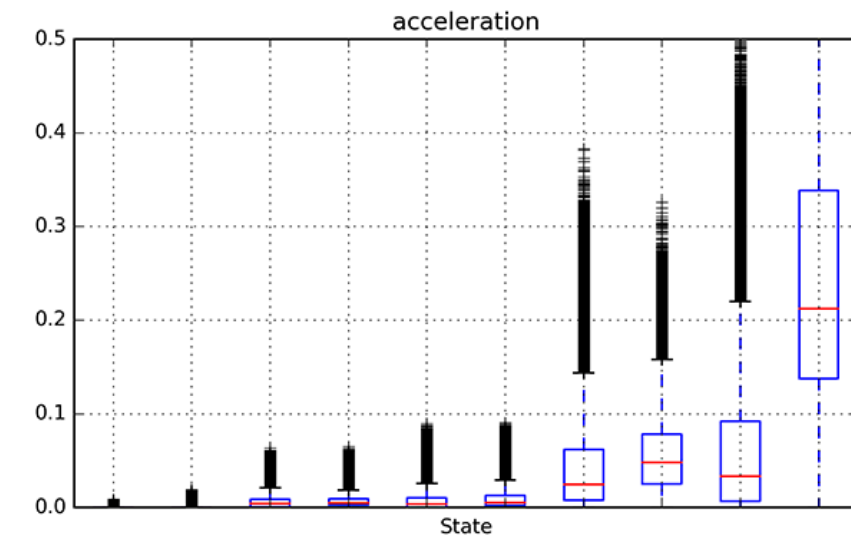
Number of states



hidden states

observations





Open Source Software for HSMM

Library	Estimation	Multi-variate observations
R package hsmm https://CRAN.R-project.org/package=hsmm	Maximum Likelihood	No
R package mhsmm https://CRAN.R-project.org/package=mhsmm	Maximum Likelihood	Yes
Python library pyhsmm https://github.com/mattjj/pyhsmm	Bayesian	Yes

Reflections on HSMM

- Lower dependence on lab calibration studies needed for cut-points approach
- Communication challenge: Community wants “time spent in MVPA”, which is not exactly what HSMM gives
- Large dataset: Train in representative subsample and then apply to remaining data?
- Multivariate: Ease of interpretation will decrease with more input variables



Questions?

VINCENTVANHEES@GMAIL.COM

WWW.MOVEMENTDATA.NL

A solid orange horizontal bar at the bottom of the slide.