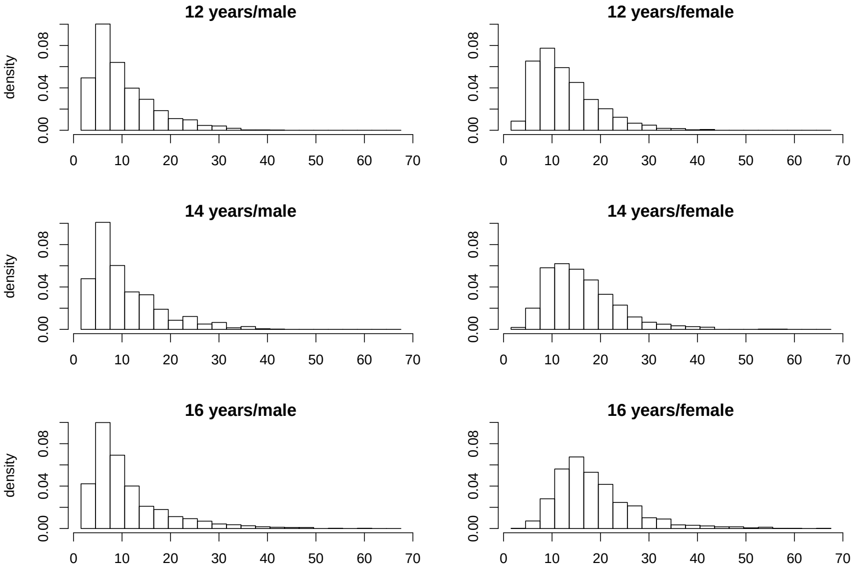
Modelling fat mass as a function of weekly physical activity profiles measured by accelerometers

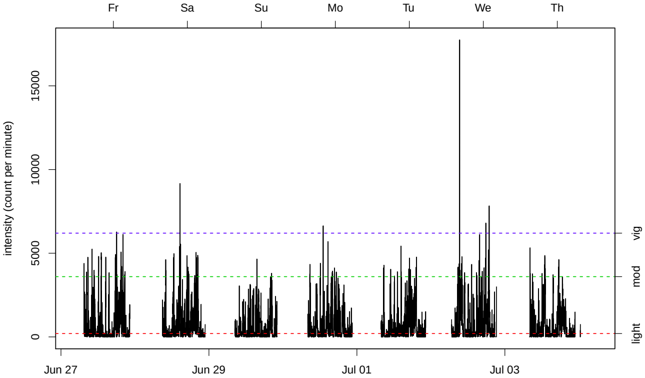
Nicole H Augustin[1], Calum Mattocks[2], Ashley Cooper[2], Andy Ness[3] and Julian Faraway[1]
[1] Department of Mathematical Sciences, Bath; [2]School for Policy Studies, Bristol; [3] Department of Oral and Dental Science, Bristol.

- The data: Avon Longitudinal Study of Parents and Children (ALSPAC);
  - physical activity and fat mass from 2003-2008 at ages 12 (n = 5500), 14 (n = 3800) and 16 (n = 2000)
- Methods and results on functional data analysis of fat mass and physical activity profiles.

Scalar response: Fat mass at ages 12, 14 and 16

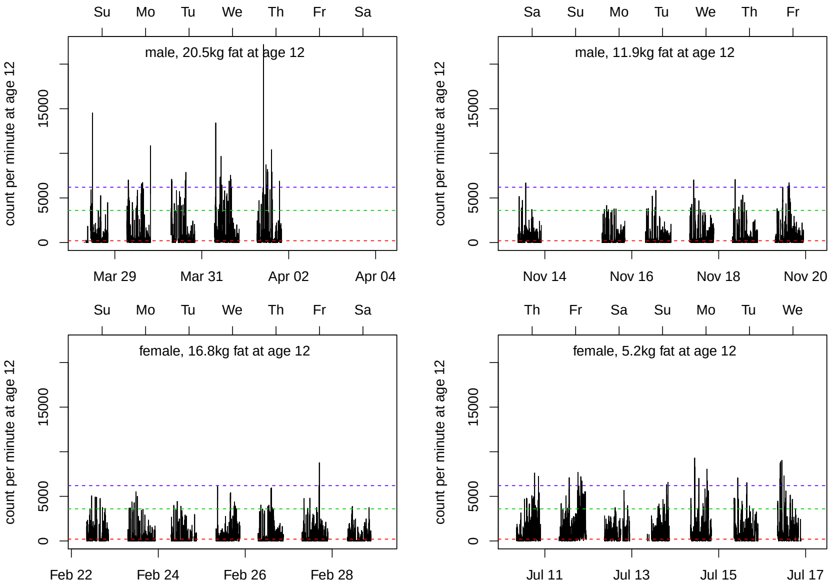


Functional predictor: Physical activity at ages 12, 14 and 16



- time series of 10080 measurements by minute over 7 days;
- after some pre-processing;
- cutpoints for moderate/vigorous activity estimated from a calibration study (Mattocks et al, 2007);
- mean hours worn per day: 11 hours (SD 4.9 hours).

Physical activity of four individuals at age 12



Objectives

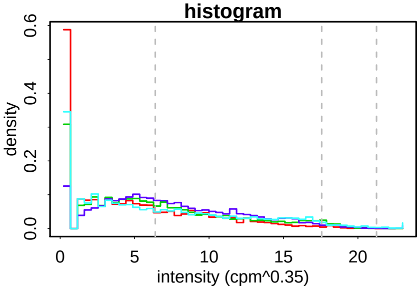
- Scientific question: What is the relationship between physical activity profile and fat mass?
- So far only single summary statistics of physical activity profiles are used,
  - `total activity`: average accelerometer counts / minute;
  - `MVPA`: average minutes of moderate to vigorous physical activity / day;
  - `average sedentary behaviour`: average minutes per day spent in sedentary activity.
- Problems: waste of information - ignore pattern and intensity distribution; dependence; cut-point dependent.

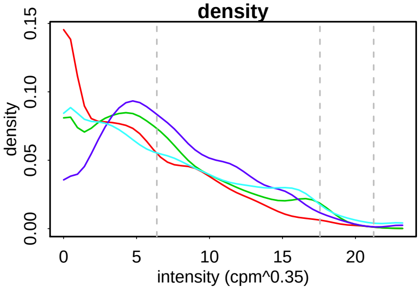
Aim: develop a cut-point independent statistical tool to explore the relationship between physical activity and fat mass.

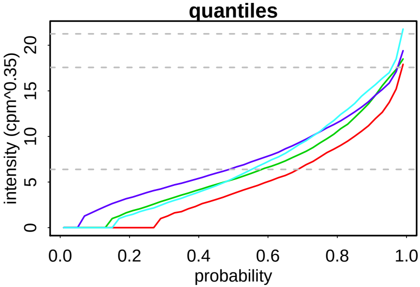
Requirement: a functional summary of the physical activity measurements

- cannot compare individuals using profiles directly;
- need to reduce the dimension of data;
- userfriendly - easy to interpret.

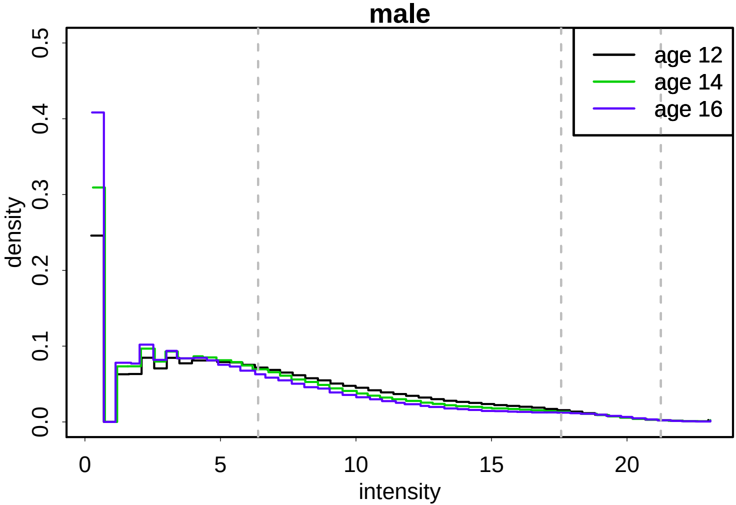
Different functional summaries



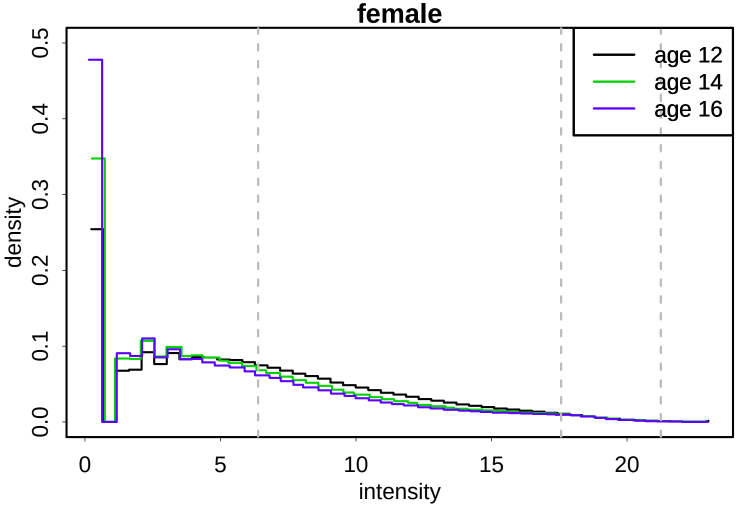




Median histogram



Median histogram



Model fat mass as a function of the accelerometer profile

- response y_{ik} total fat mass for individual i at age k;
- vector x_{ik} is the accelerometer profile, with 10080 entries;
- $z_{ik}(x)$ is the histogram with some given number of mid-points x_j;

Start with a linear model:

$$log(y_{ik}) = \alpha + \sum_j \beta_j z_{ik}(x_j) + \sum_l \gamma_l \texttt{confounder}_{lik} + \epsilon_{ik}.$$

with $\epsilon_{ik} \sim N(0, \sigma^2)$ and confounders `sex`, `height`, `height`^{2}.

A generalised regression of scalars on functions (Ramsay and Silverman, 2005)

Let the β_j vary smoothly, where $\beta_j = f(x_j)$:

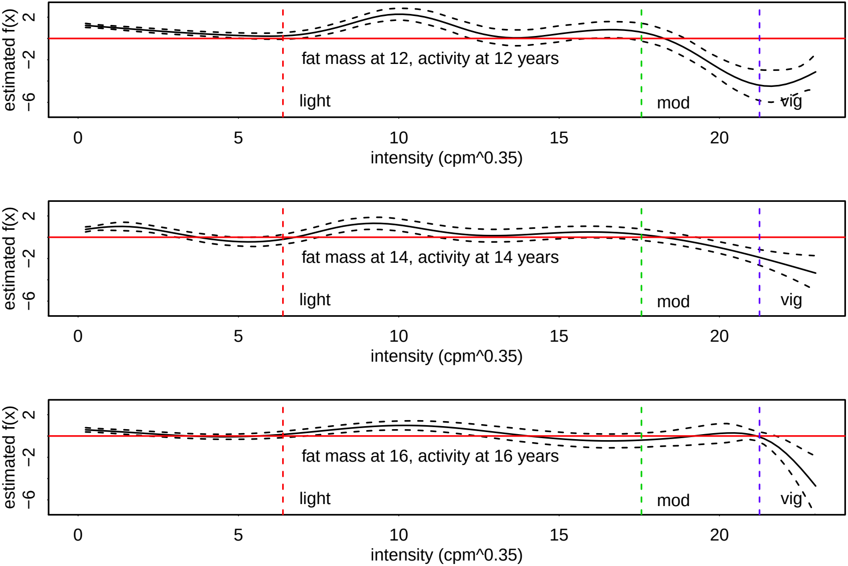
$$log(y_{ik}) = \alpha + \sum_j f(x_j) z_{ik}(x_j) + \sum_l \gamma_l \texttt{confounder}_{lik} + \epsilon_{ik}.$$

- $f(x)$ is an unknown 'coefficient' function to be estimated;
- $f(x)$ is represented using an adaptive smooth with a P-spline basis (Eilers & Marx, 1996);
- based on B-spline basis functions and discrete penalties on the basis coefficient;
- adaptive: terms in penalty have different weights depending on j;
- smoothness parameters λ_j are multiplied by weights.

Parameter estimation - nested iteration scheme (Wood, 2011)

- Outer iteration is approximate restricted maximum likelihood (REML) estimation of smoothness parameters
  - Inner iteration is penalised iterative re-weighted least squares (PIRLS) to find all other parameters (coefficients of basis functions, and coefficients of linear terms).
- Scheme is implemented in the gam() function of the mgcv R package.

Results



cross-sectional analysis, fitting separate models for three ages.

Further work

- longitudinal model;
- other types of summary functions with regards to temporal activity pattern;
- other health outcomes.

References

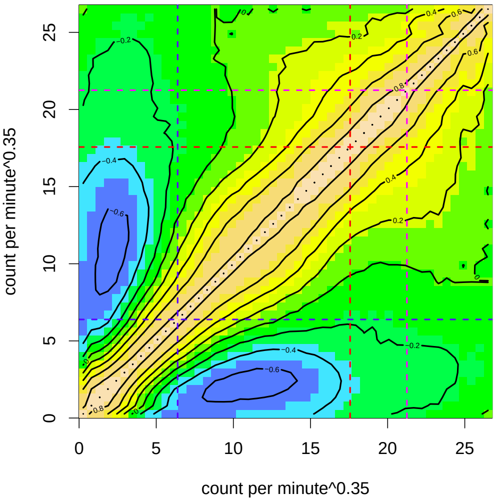
Mattocks, C., Ness, A., Leary, S., Tilling, K., Blair, S., Shield, J., Deere, K., Saunders, J., Kirkby, J., Smith, G. D., Wells, J., Wareham, N., Reilly, J., and Riddoch, C. (2008). Use of accelerometers in a large field based study of children: Protocols, design issues, and effects on precision. *Journal or Physical Activity and Health*, 5:S94–S107.

Mattocks, C, Griffiths, A, Tilling, K, Blair, SN, Leary, S, Deere, K, Davey Smith, G, Pate, RR, Ness, A, Riddoch, C (submitted). Prospective associations between objectively measured physical activity and fat mass in adolescents.

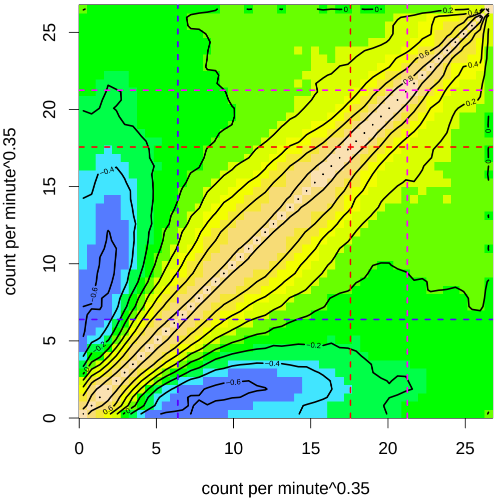
Ramsay, J.O. and Silverman, B.W. 2005. Functional Data Analysis. Springer Verlag.

Wood, SN (2011). Fast stable REML and ML estimation of semiparametric GLMs. *Journal of the Royal Statistical Society, Series B*, 73, pp. 3-36.

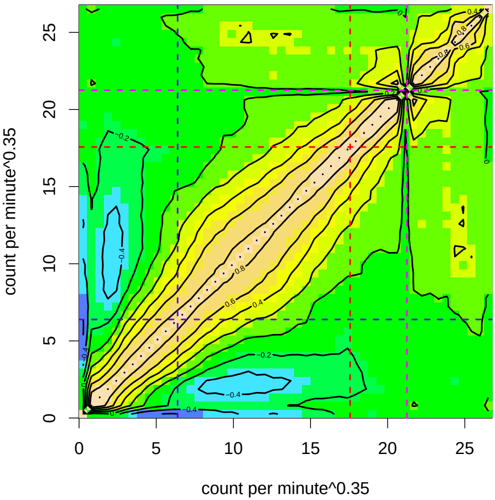
Correlation matrix of histograms at age 12



Correlation matrix of histograms at age 14



Correlation matrix of histograms at age 16



Protocol for pre-processing of activity profiles

1. replace any sequence with more than 10 zeros by missing values;
2. exclude days if:
   - mean count < 150;
   - mean count > 3 SD above overall mean (prior to exclusions);
   - monitor was worn < 10h;
3. Exclude weekly profiles if < 3 valid days were observed.