Functional Data Analysis

An introduction

David Wooff

d.a.wooff@durham.ac.uk

University of Durham

Main steps in FDA

- Collect, clean, and organize the raw data.
- Convert the data to functional form.
- Explore the data through plots and summary statistics
- *Register* the data, if necessary, so that important features occur at the same argument values.
- Carry out exploratory analysis, such as functional principal components analysis
- Construct models, if appropriate
- Evaluate model performance

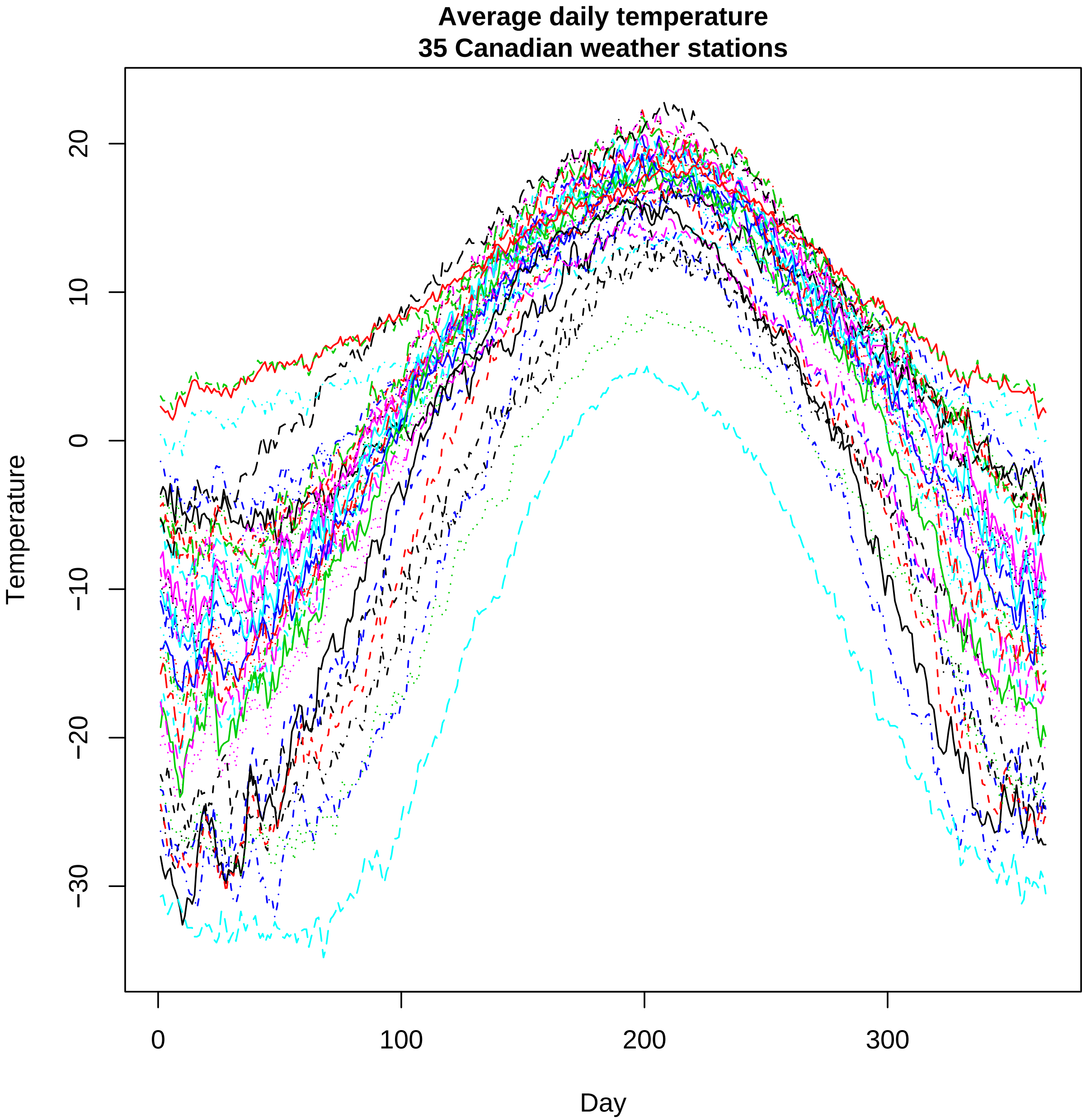
Durham University

Example: Weather data

- Daily average temperature computed for $N = 365$ consecutive days at each of 35 Canadian weather stations
- Are there common patterns of interest?
- Is there variation in *amplitude* (peakedness)?
- Is there variation in *phase* (timing of features)?

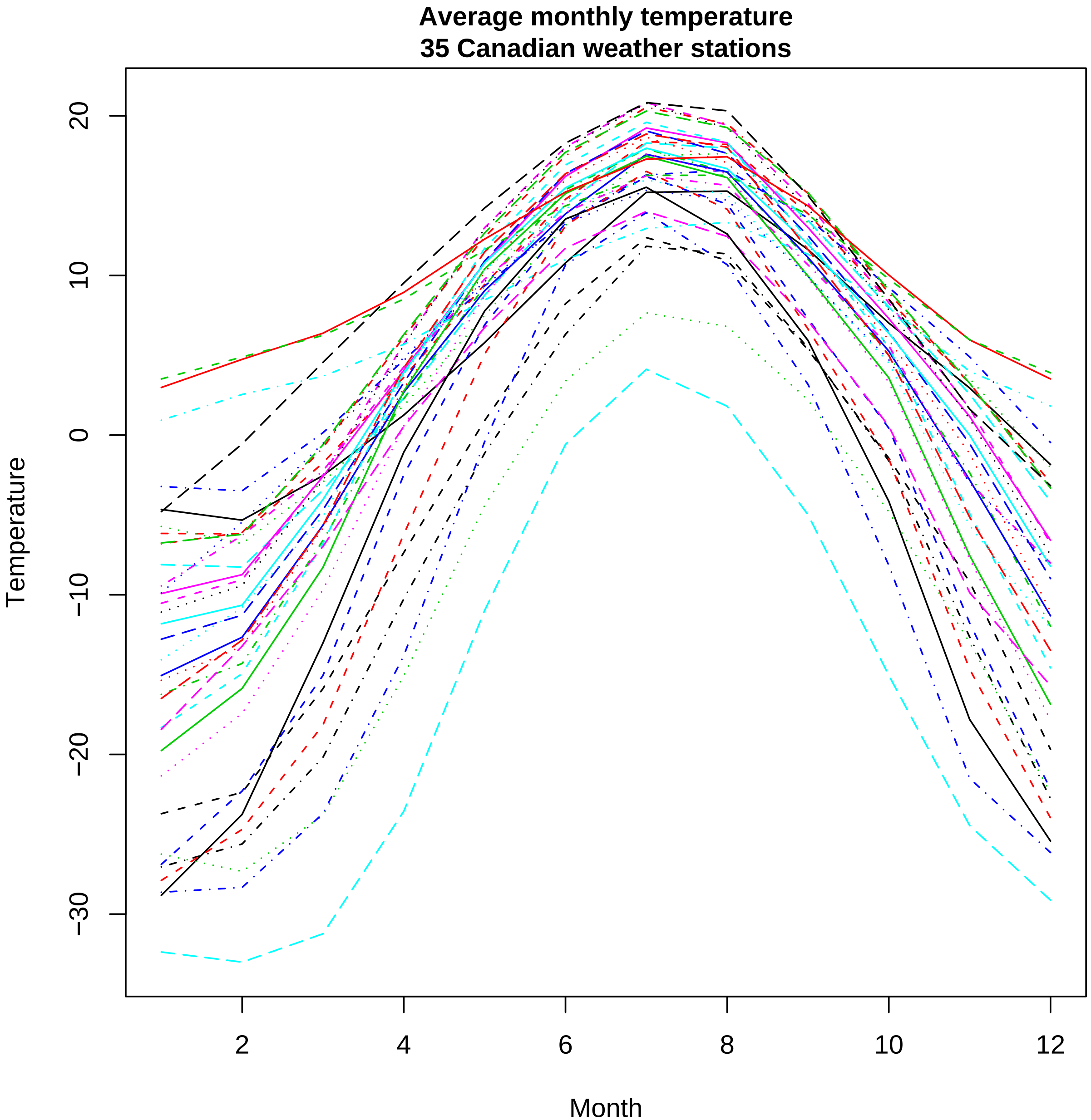
Durham University

Raw data: Daily mean temperatures at 35 Canadian weather stations



Durham University

Averaged data: Monthly mean of daily mean temperatures at 35 Canadian weather stations



Durham University

Summary of visual features

- Clear that temperatures rise in the summer, fall in the winter
- Differences in level: some very cold places in arctic Canada
- Strong differences in amplitude.
  - Coastal stations display little amplitude: cool winters and summers.
  - Continental stations show high amplitude/peakedness: cold winters and hot summers.
- Some variation in phase: peak for some (coastal?) stations is after the peak for others (continental?)

Durham University

Basis construction

- Main idea is that we take each observed series $x_i(t)$ and approximate it by $\hat{x}_i(t)$ chosen from the same functional family.
- We represent a function x by a linear expansion

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

in terms of K known basis functions $\phi_k(t)$.

- $c_1, \ldots, c_K$ are coefficients to be estimated.
- Let n be the sample size, i.e. the number of observations in each series.

Durham University

Basis construction

- We hope that small K leads to a reasonable fit and the capture of essential characteristics, balanced by large degrees of freedom to allow computation of CIs.
- Once we have the approximation, we have

  $$x(t) = \hat{x}(t) + \epsilon(t),$$

  and observed residual series

  $$\hat{\epsilon}_i(t) = x_i(t) - \hat{x}_i(t),$$

  from which we may obtain a standard measure of quality of fit for an individual series:

  $$s_i^2 = \frac{1}{n-K} \sum_{j=1}^{n} \hat{\epsilon}_i(t_j)^2.$$

Basis choice

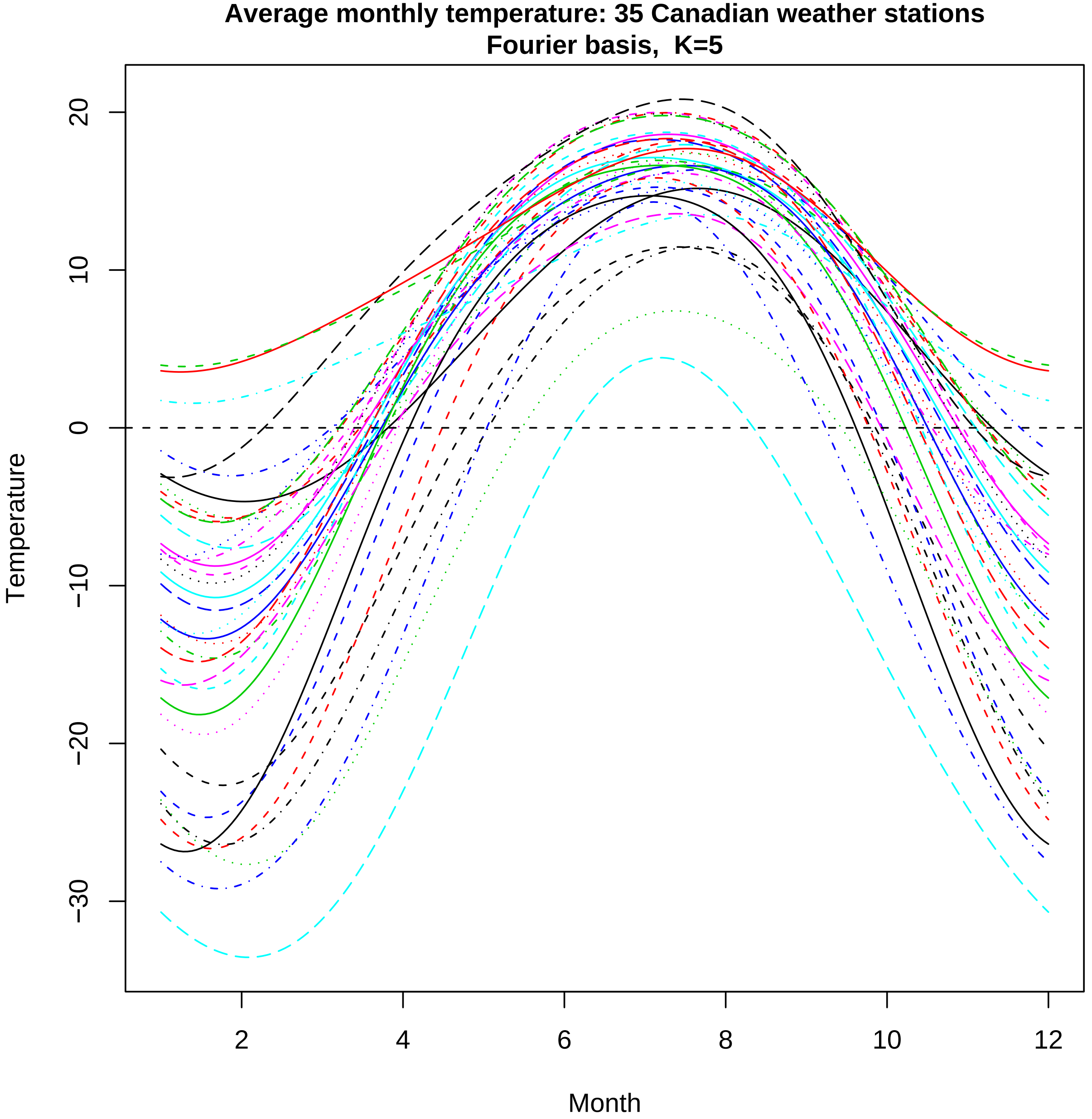
- Several possibilities: Bsplines, exponential, polynomial ($\phi_k = t^k$). Bsplines mostly used for aperiodic data.
- Reasonable to assume that the weather data is periodic
- Reasonable choice is Fourier basis system:

  $$\hat{x}(t) = c_1 + c_2 \sin \omega t + c_3 \cos \omega t + c_4 \sin 2\omega t + c_5 \cos 2\omega t + \ldots,$$

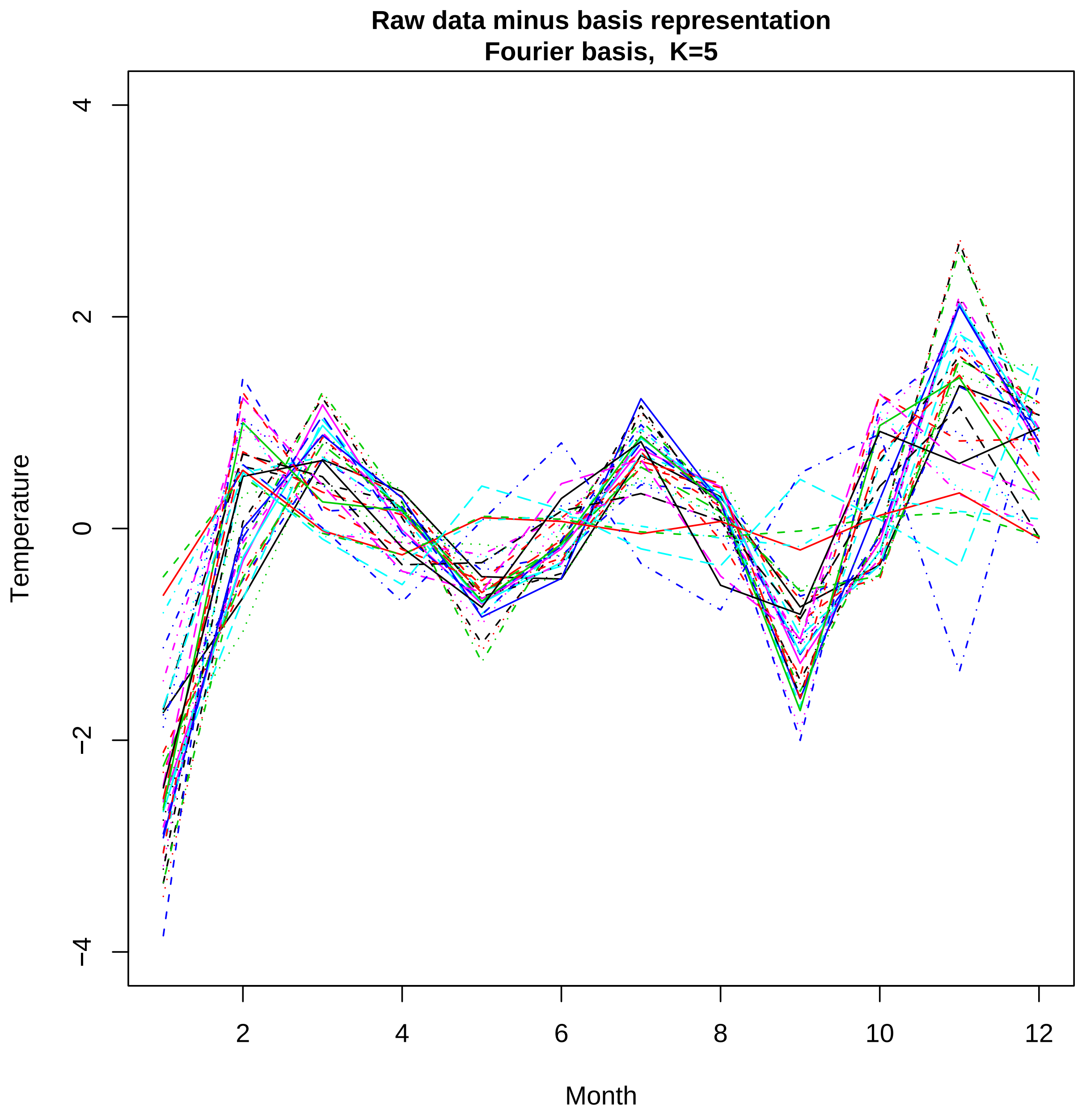
  where $\omega = 2\pi/T$, here $T = 11$
- Here we choose $K = 5$; for a Fourier basis system we choose an odd number to capture variation in *phase*, i.e. we require $sine$, $cosine$ pairs
- Various ways of estimating $c_1, \ldots, c_K$: the simplest is ordinary least squares (used for this example), but other approaches - e.g. roughness penalty - may be superior.

Durham University

The basis curves, $\hat{x}(t)$ for the weather data



Residual plots, $(\hat{x}_i(t) - x_i(t))$, $i = 1, \ldots, 35$, for $K = 5$.



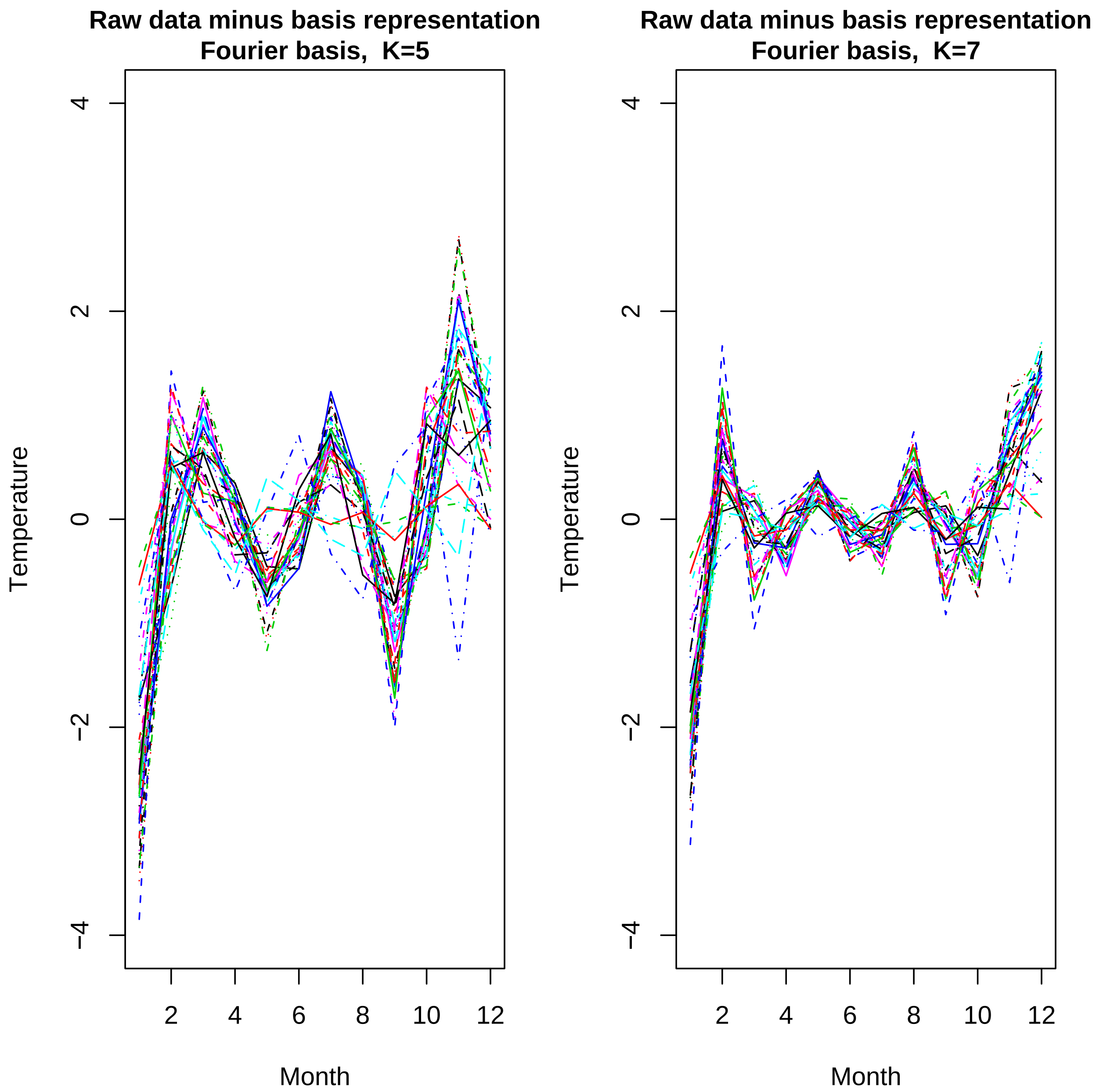
Durham University

Residual plots

- Estimated curves appear plausible
- Should also explore the fitted curves: plot raw residuals $\hat{\epsilon}_i(t) = x_i(t) - \hat{x}_i(t)$.
- Periodicity in the residuals is probably an artefact of the basis representation.
- Do we get a better answer if we choose more basis functions?

Durham University

Residual plots for $K = 5$ and $K = 7$



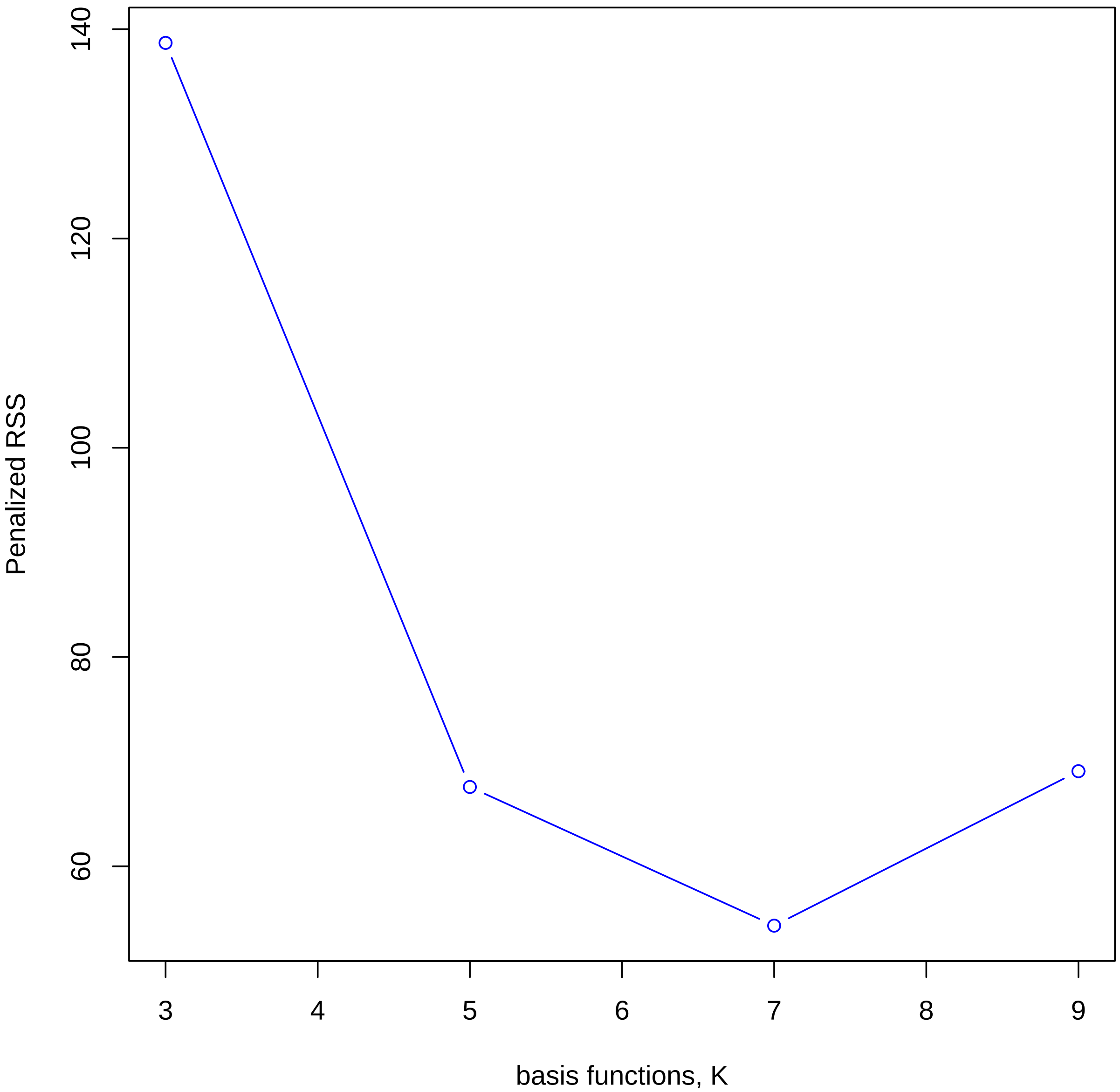
Durham University

Residual plots

- Raw residuals decrease as we increase K (of course), but the model is more complicated.
- Simple way of assessing the number of basis functions to use is to examine $s_i^2 = \frac{1}{n-K}\sum(\hat{x}_i(t) - x_i(t))^2$
- For example, we can plot $\sum s_i^2$ versus K - note that s_i^2 is penalized by the number of basis functions used
- Analagous to the usual measure of variability in LS regression, $\frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

Durham University

Plots of $\sum s_i^2$ versus K



Durham University

Number of basis functions

- The plot suggests that there is a big advantage in using $K = 5$ rather than $K = 3$
- Not much extra advantage in using $K = 7$: we are starting to lose degrees of freedom
- Pragmatic choice seems to be $K = 5$
- This is a black art!

Durham University

Exploring derivatives

- Derivatives of the fitted curves can be explored and can be interpreted
- First derivative may suggest common *landmarks*
- Plots of derivatives may suggest certain models, eg the differential equation

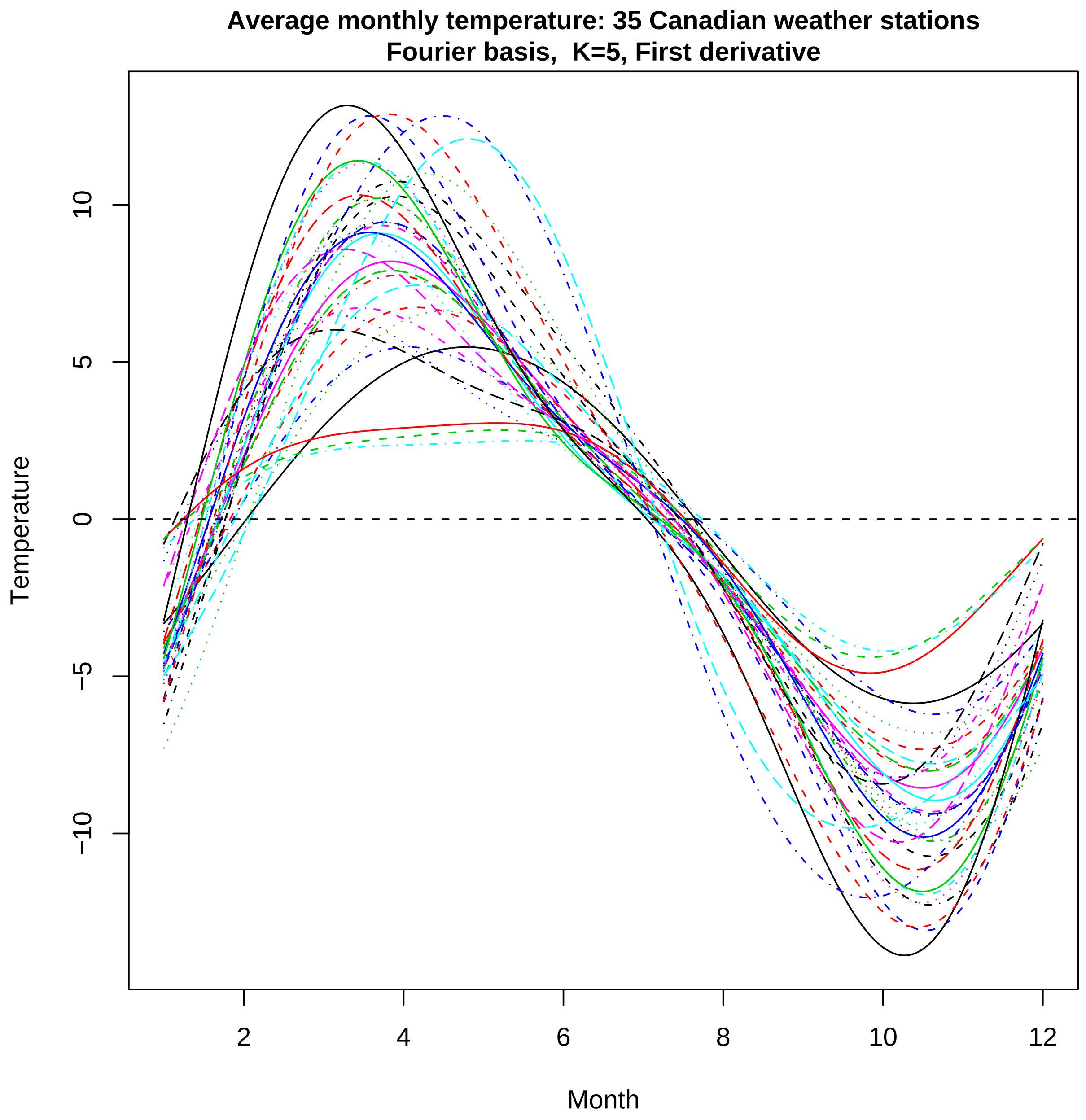
$$\Delta x = -\alpha(x - b_1)$$

is consistent with an exponential model,

$$x(t) = b_1 + b_2 e^{\alpha t},$$

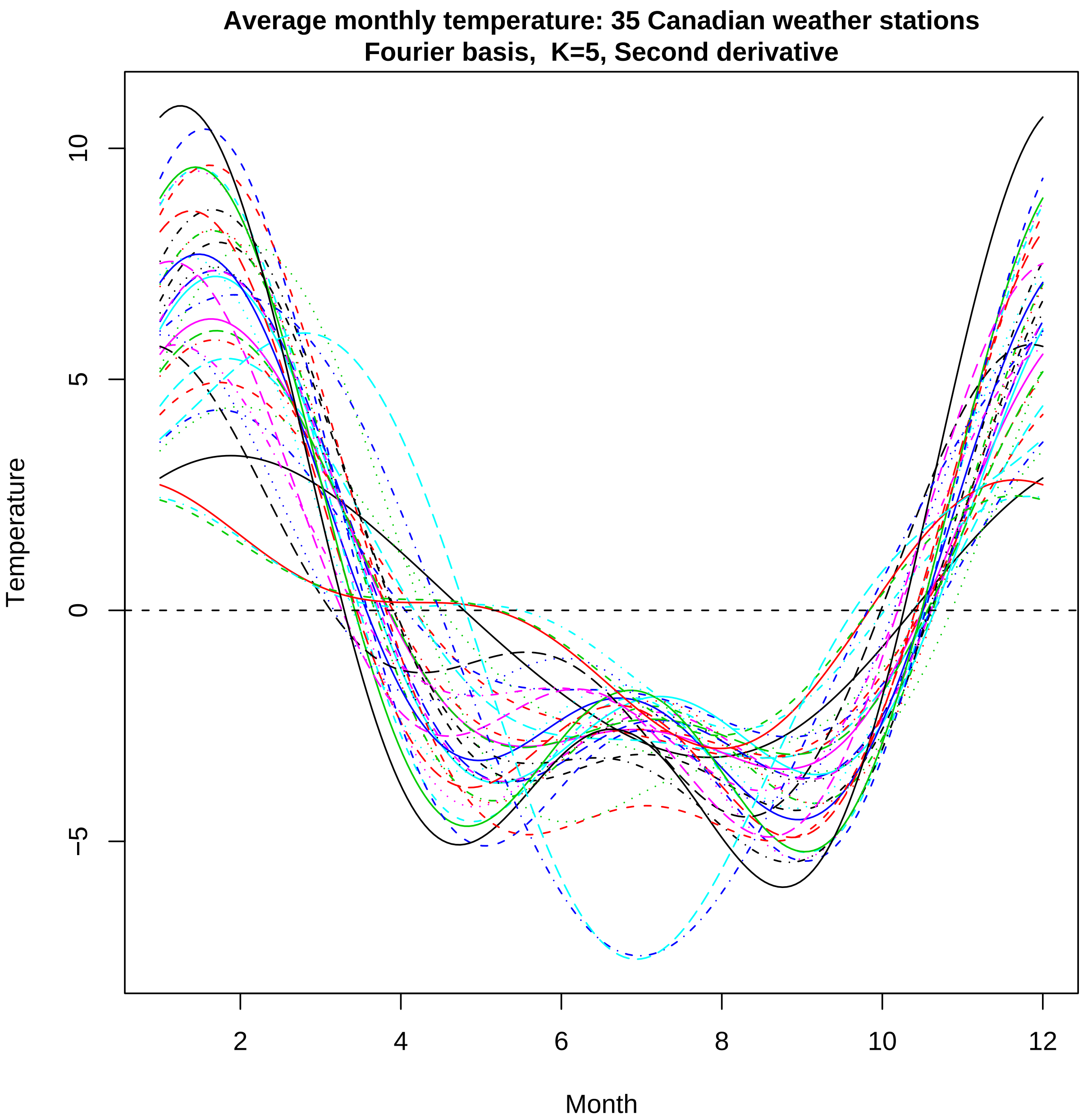
so a plot showing linear association between $x(t)$ and $x'(t)$ suggests the exponential model.

The basis curves, $\hat{x}'(t)$



Durham University

The basis curves, $\hat{x}''(t)$



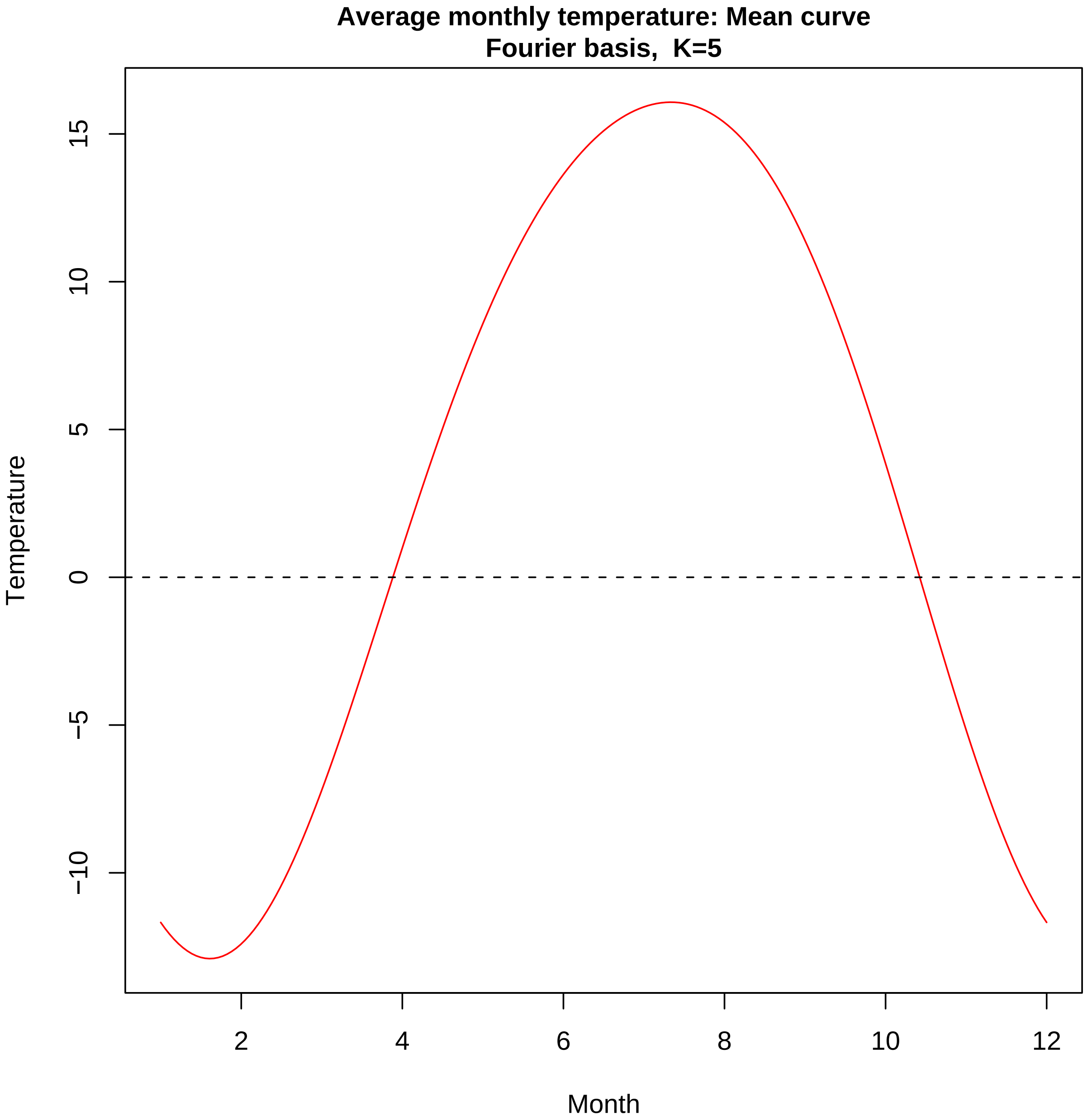
Durham University

Derivatives plots

- Plot of first derivative shows the locations of winter minima and summer maxima at $\hat{x}'(t) = 0$.
- These don't quite align, indicating that the change points differ according to location
- Plot of second derivative can highlight more subtle differences - e.g. the July shape is peaked for most locations, but troughed for others.
- We may need to be careful when interpreting such features, they may simply be artefacts of the fitting process.
- Avoid post-hoc rationalization

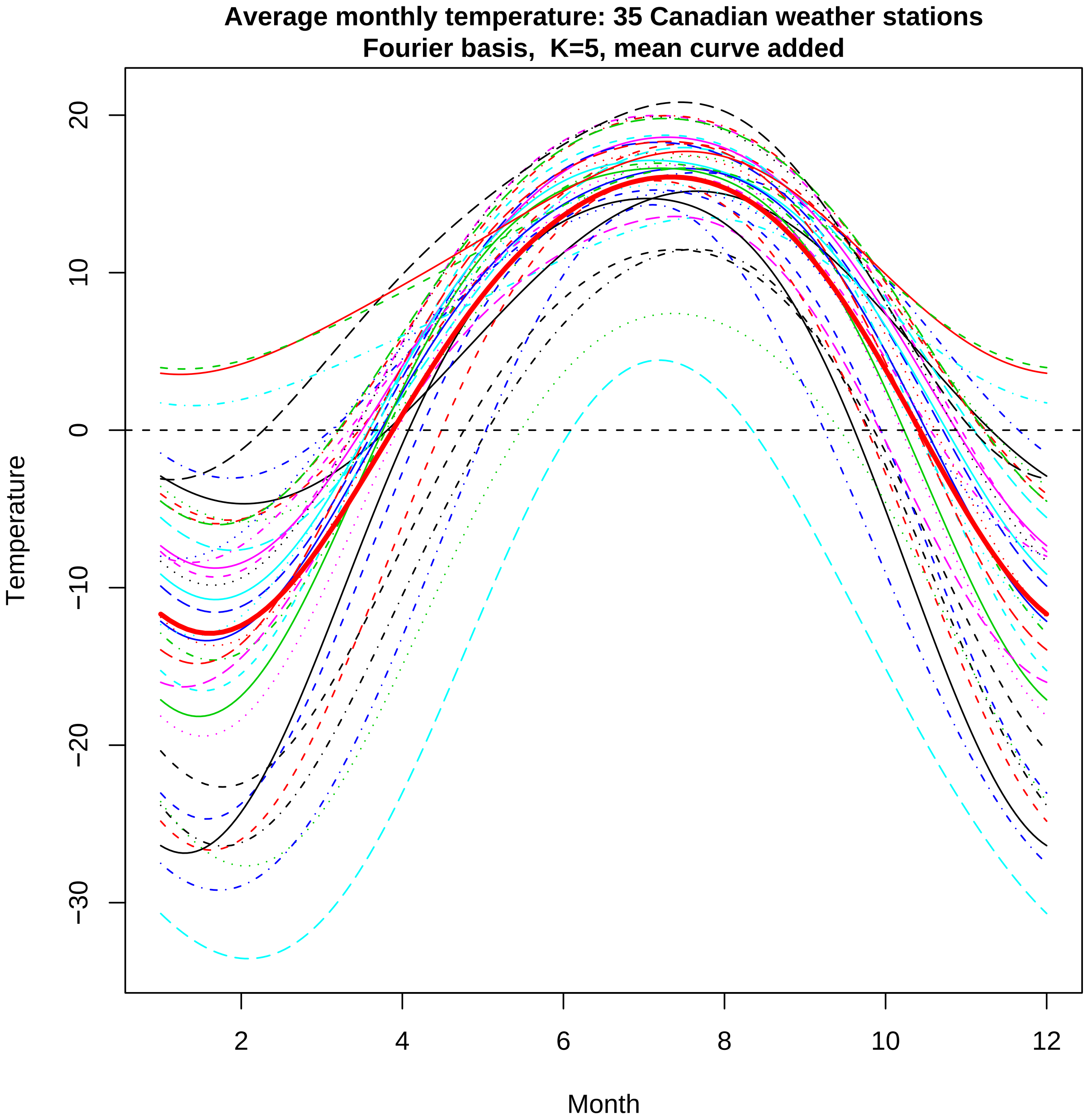
Durham University

Plotting the mean curve, $\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} \hat{x}_i(t)$

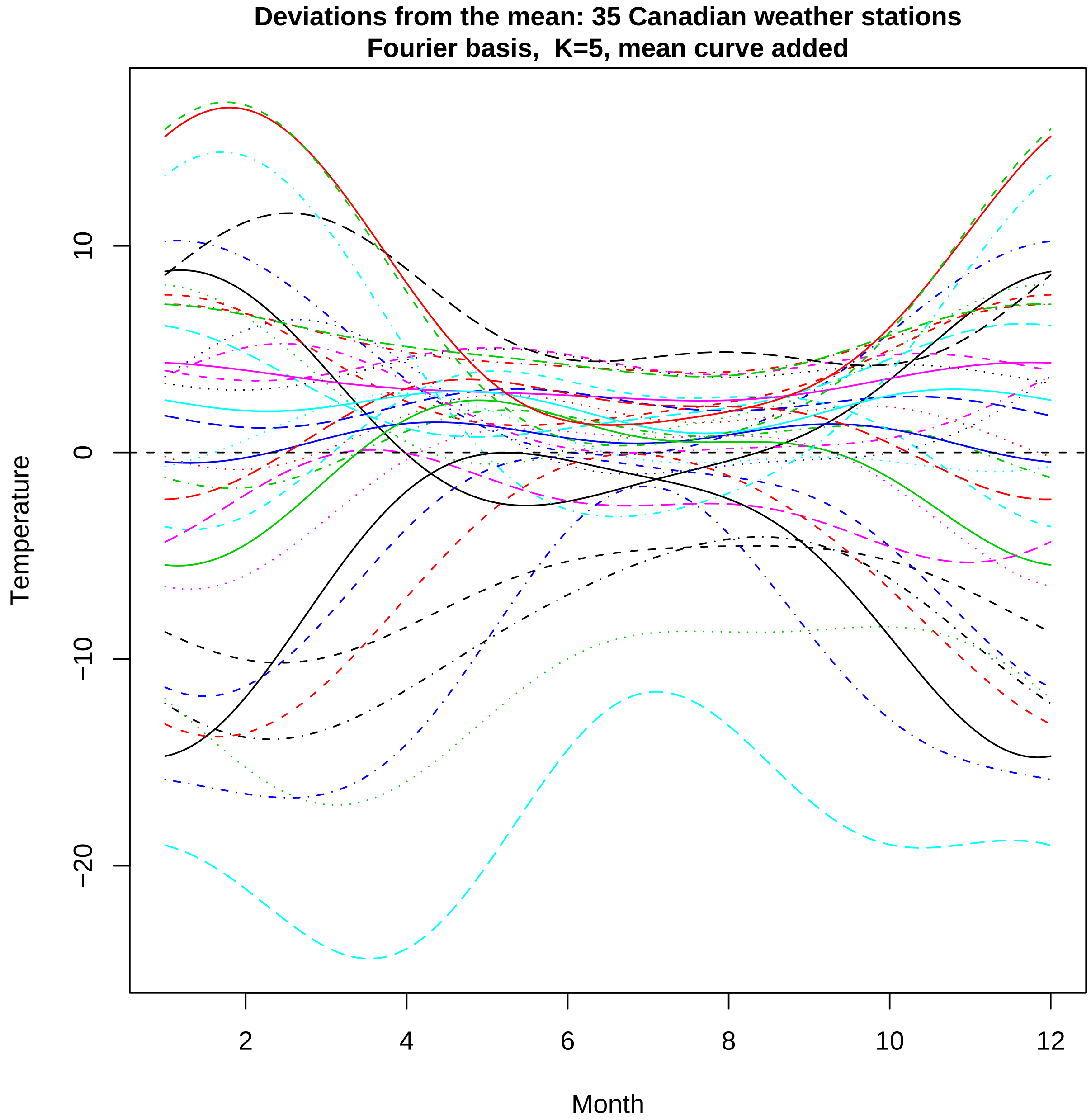


Durham University

Superimposing the mean curve on the estimated curves



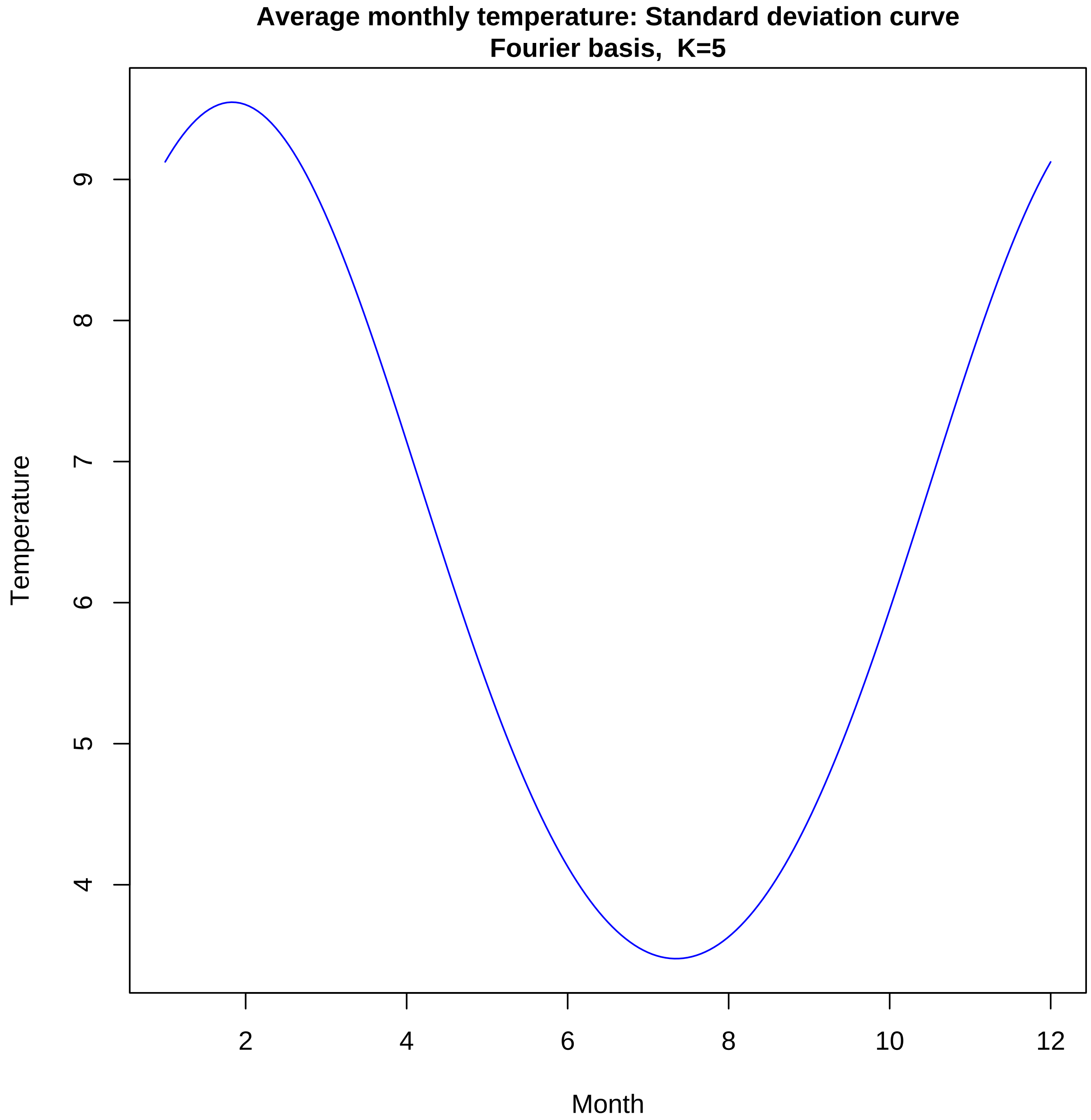
The deviations from the mean curve: shows quite a lot of variation beyond the mean



Durham University

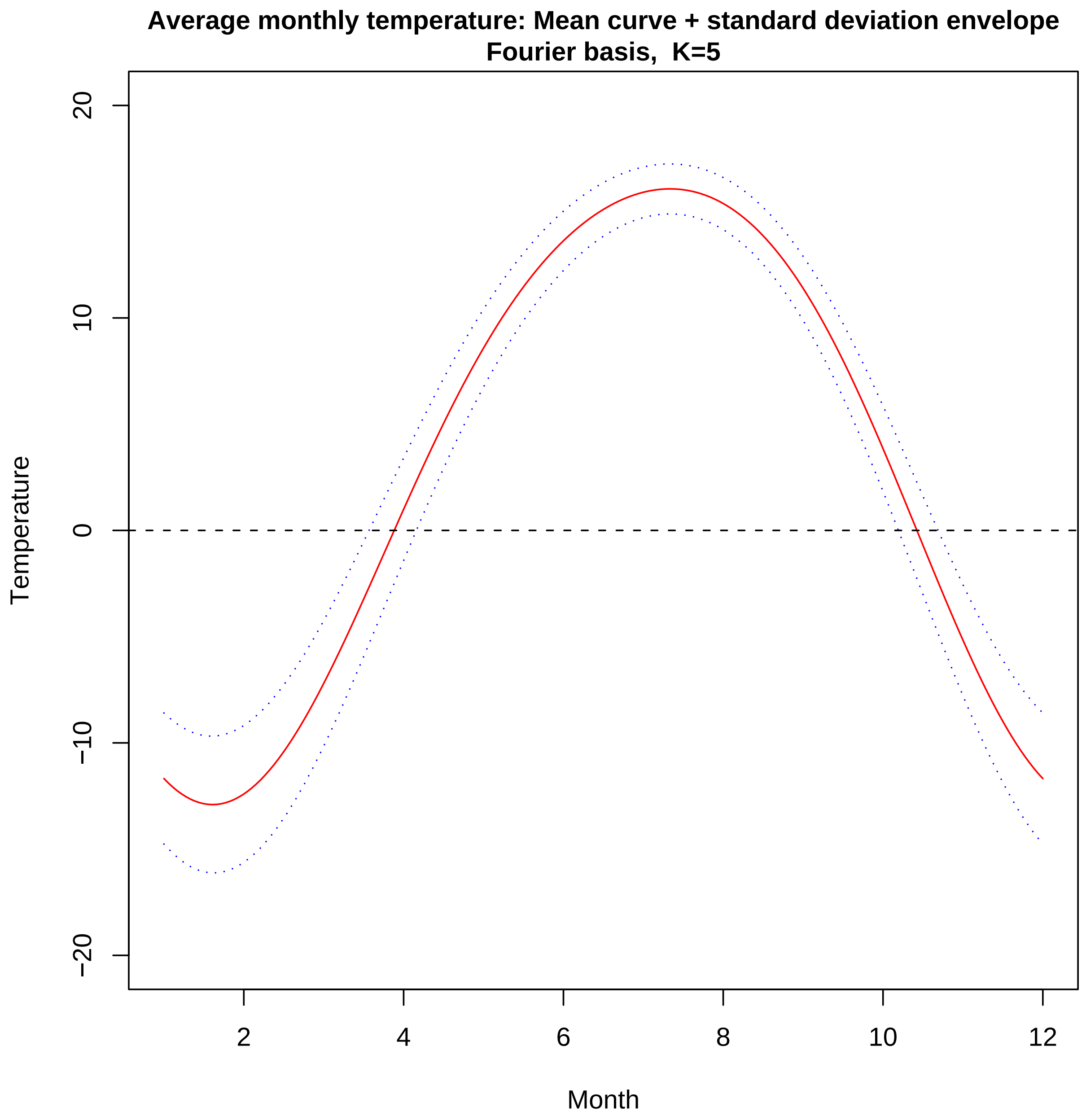
Plotting the standard deviation curve,

$$s_X(t) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{x}_i(t) - \bar{x}(t))^2}$$



Mean curve with ad hoc SD envelope based on

$$\bar{x}(t) \pm \frac{1}{\sqrt{N}} s_X(t)$$



Durham University

Interpretation

- Variation across the series is highest in the winter months, and lowest in the summer
- Can plot an ad-hoc confidence envelope for the underlying mean curve, assuming these 35 stations to be a random sample from a superpopulation of weather stations.
- Prediction envelope for a new weather station would be much wider.

Durham University

Principal components analysis for functional data

- To extract the first functional PC we seek a further function $\zeta_1(t) = \int \beta(s)x(s)ds$ which in some sense maximises variation over the space of interest.
- This requires choosing weight functions $\beta(s)$.
- This is analagous to standard PCA, in which we choose weights $\beta_1, \ldots, \beta_n$ to maximise

$$var(\sum_{i=1}^{n} \beta_i x_i).$$

- Subsequent PCs are $\zeta_2, \ldots, \zeta_K$, with the extraction arranged so that the PCs are orthogonal:

$$\int \zeta_i(s)\zeta_j(s)ds = 0, \quad \forall i \neq j.$$

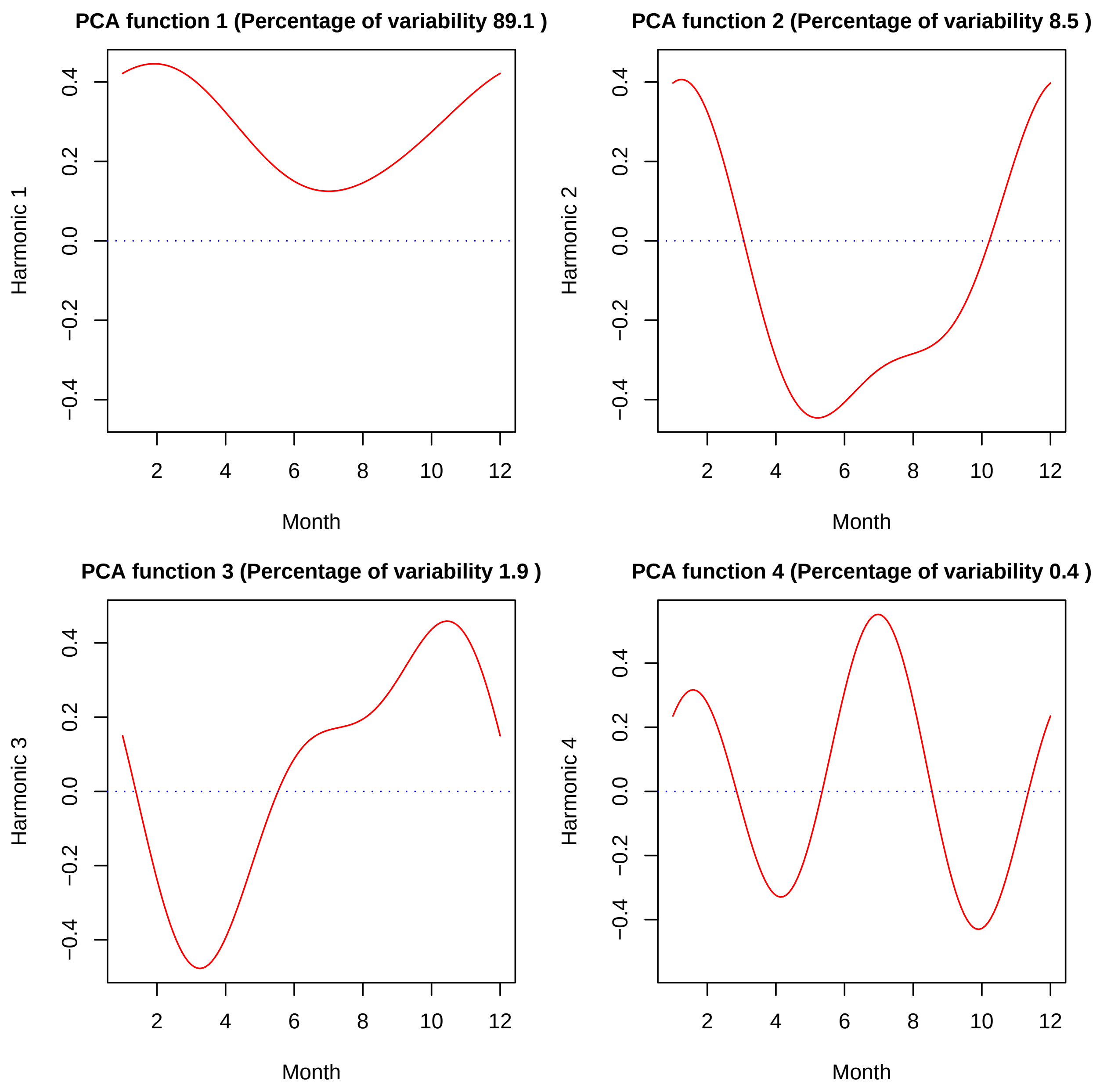
Durham University

Principal components estimation

- The estimation of the PCs is usually carried out on centred data, and so the components may be considered as perturbations of the mean curve.
- Estimation procedure is quite complicated: see Ramsay & Silverman.

Durham University

Monthly data PCA, raw components shown as weights



Durham University

The first PC for the weather data

- $\zeta_1(t)$ explains 89.1% of the variation. It contrasts winter months (high variation) with summer months (low variation), i.e. the weights placed on winter measurements are higher than than the weights for summer measurements, and so lead to higher variability.
- Scores can be computed for each data series (weather station), as $\zeta_{1i}(t) = \int \beta(s)x_i(s)ds$.
- Weather stations with high positive values of $\zeta_{1i}(t)$ will have much warmer than average winters and warmer than average summers.
- Weather stations with high negative values of $\zeta_{1i}(t)$ will have much colder than average winters and colder than average summers.

Durham
University

Subsequent PCs for the weather data

- $\zeta_2(t)$ explains 8.5% of the variation. It contrasts warm winter months with cool summer months.
- Weather stations with high positive values of $\zeta_{2i}(t)$ will have warmer than average winters and colder than average summers.
- Weather stations with high negative values of $\zeta_{2i}(t)$ will have colder than average winters and warmer than average summers (for example, for mid-continental stations).
- The remaining PCs appear consonant with random variation and explain little of the variation.

Durham University

Re-adding the mean curve to the PCs

- An alternative visualization is obtained by plotting

$$\bar{x}(t) \pm \gamma\zeta_j$$

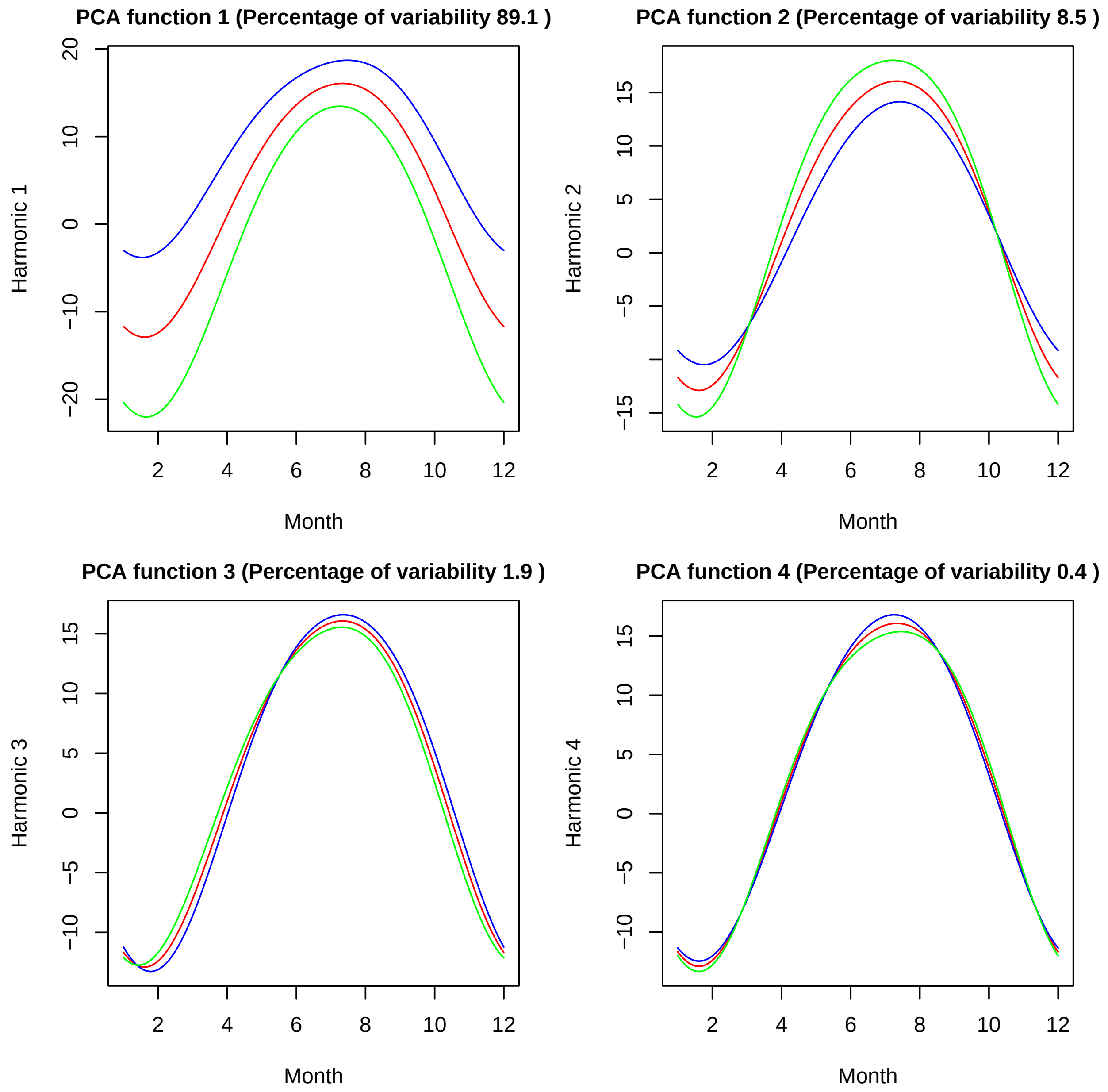
  for some appropriate value of γ.

- That is, we add the average to the PCs and so can see the implication of high positive and high negative values for each $\zeta(t)$.

- γ can chosen by inspection, to distinguish sufficiently between the curves; alternatively Ramsay et al suggest a semi-automatic procedure for choosing γ.

Durham University

Monthly data PCA: plotting components as perturbations of the mean



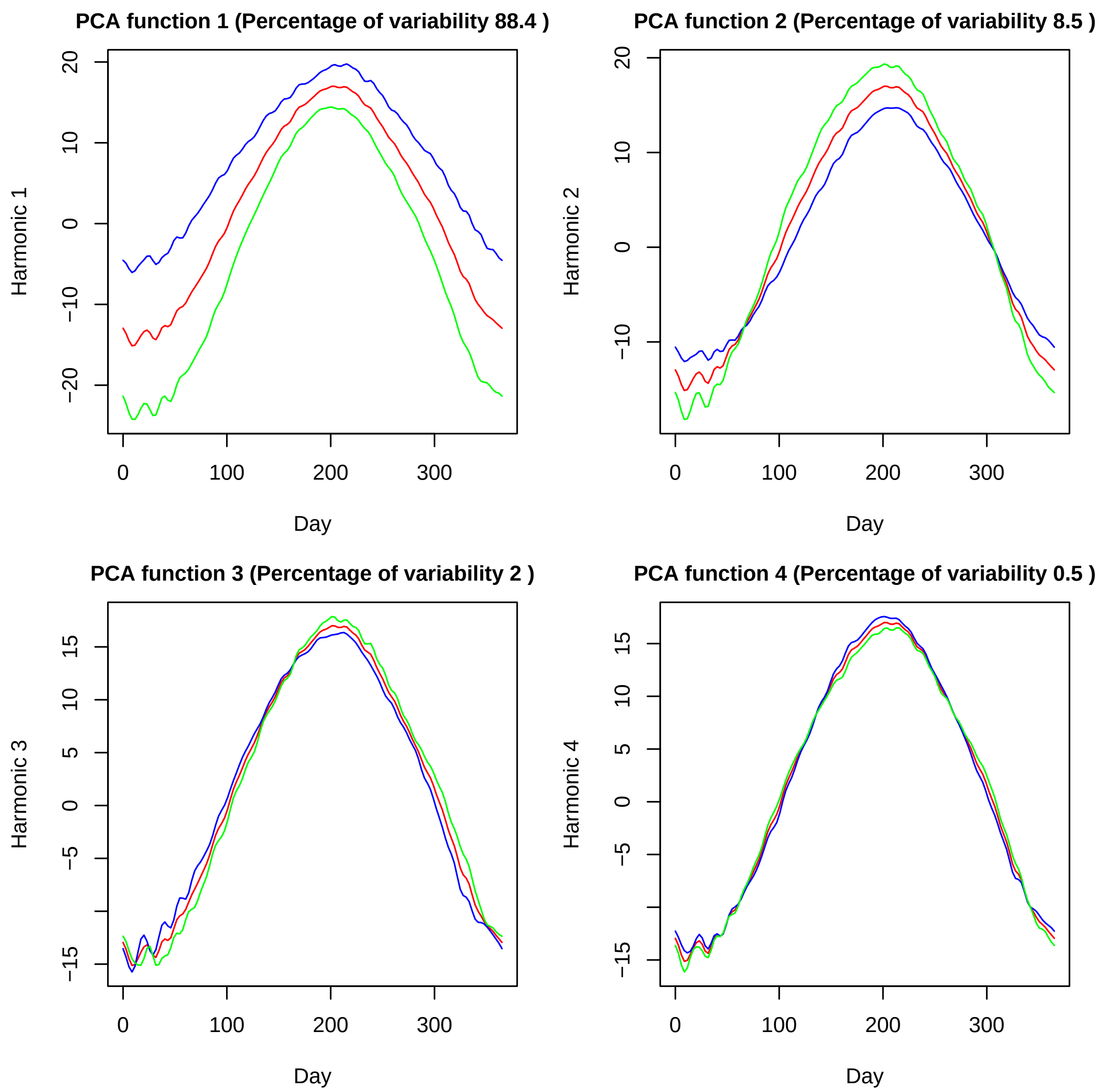
Durham University

Interpreting the remaining PCs

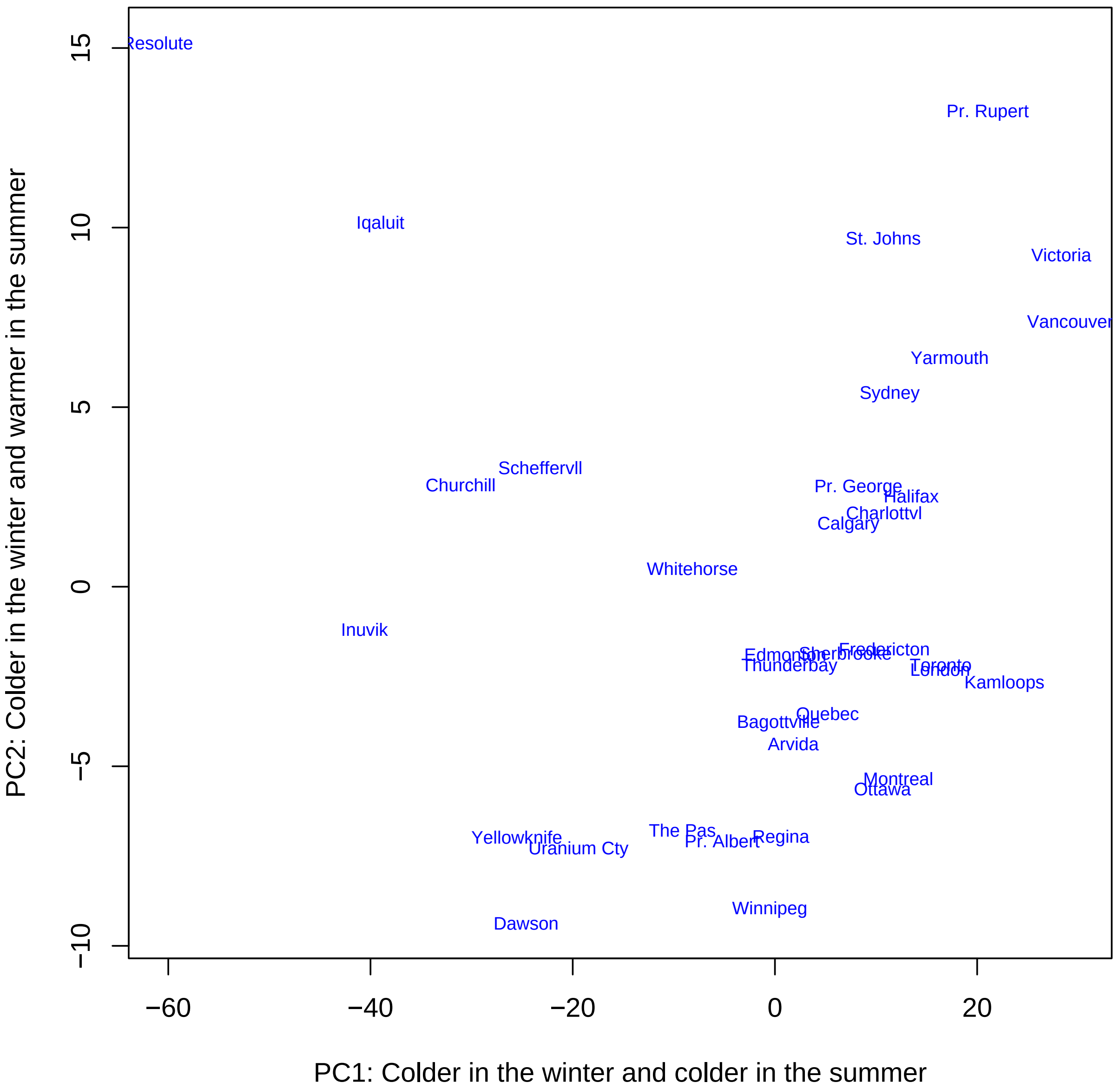
- Using this visualization, we might now interpret $\zeta_3(t)$ as a slight time shift component: high values have slightly later summers and winters, and vice-versa for low values.
- Still hard to interpret $\zeta_4(t)$, but might concern timing of Autumn and Spring.
- It is also possible to carry out rotations of the PCs in order to improve interpretability.
- These can be orthogonal or oblique, as desired.
- Oblique rotations lead to correlated components, which is ok in my view, but . . . . . .
- . . . . . . rotation doesn't give appreciably better answers for this data.

Durham University

Daily data PCA, $K = 65$. We don't seem to have lost much by smoothing from daily to monthly data.



Plot of PC scores of 35 weather stations. Neighbouring stations have similar PC scores



Durham University

Registration

- The weather data have a natural starting and ending position. This is not so for other data.
- After transforming to functional form, it may be necessary to *align* them prior to further exploration.
- The process of *aligning* curves according to their principal features is called landmark registration.
- Example: 68 cross-sections of ends knee bones were taken; each datum is a bitmap image.
- 12 landmarks were identified as capturing the essential features
- A smooth curve is fitted through the landmarks and standard fda methods applied

Durham University

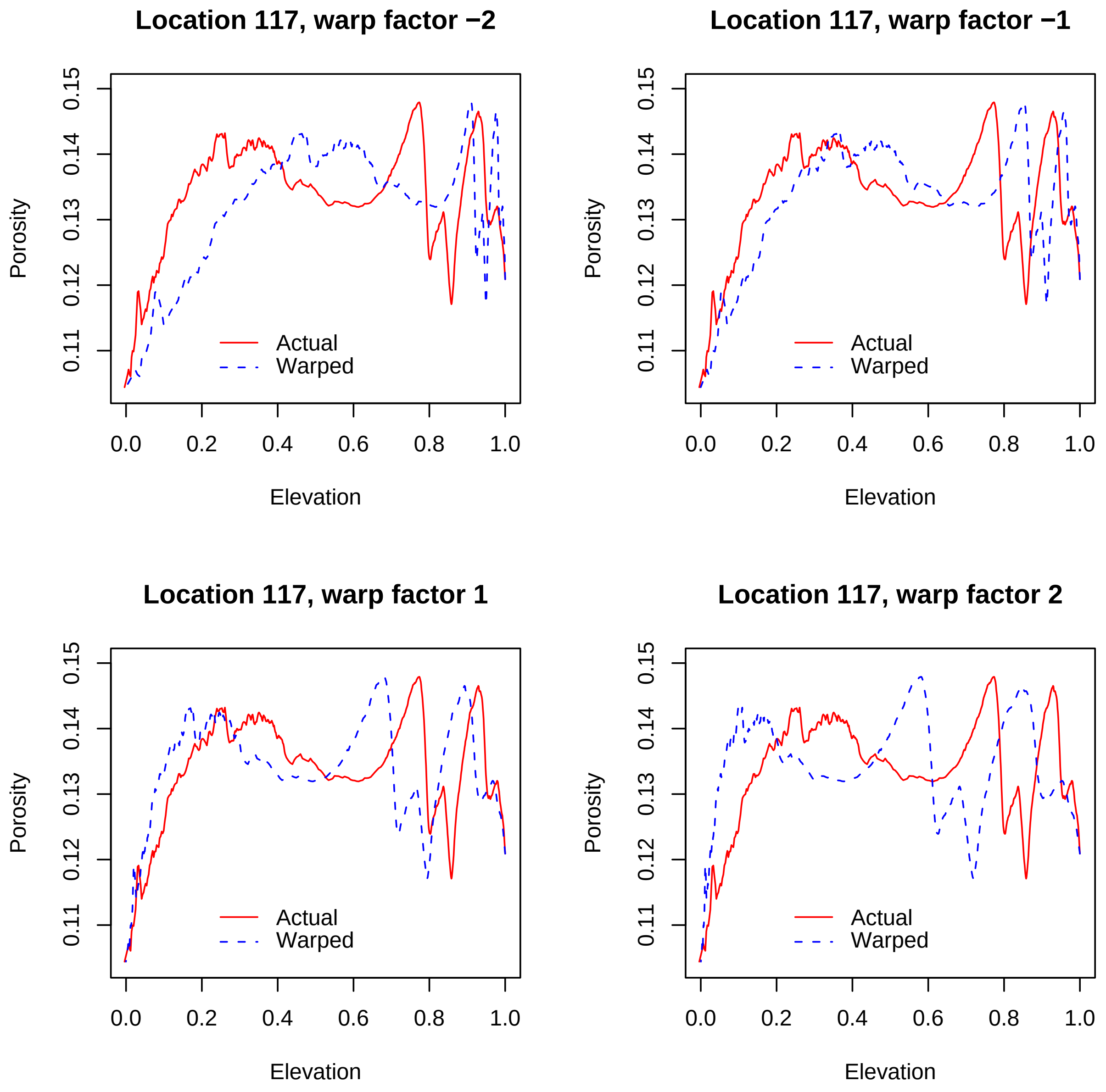
Warp functions

- One way of aligning functional objects is via a warping function.
- A warping function maps from $(0, T)$ to $(0, T)$ in the x direction, but compresses part of the range whilst stretching the remainder.
- An example with warping parameter α is:

$$h(x, \alpha) = \frac{e^{\alpha x} - 1}{e^{\alpha} - 1}, \quad \alpha \neq 0.$$

- $\alpha < 0$ deforms to the right and $\alpha > 0$ deforms to the left. We have $\lim_{\alpha \to 0} h(x, \alpha) = x$, so that a value of $\alpha = 0$ implies no warping.
- Each data object requires the warping parameter to be estimated.

Warping applied to height in order to align oil reservoir porosities, $h(x,\alpha) = (e^{\alpha x} - 1)/(e^{\alpha} - 1)$



Durham University

Other techniques

- Canonical correlation: which modes of variability in two sets of curves are most closely associated?
- Discriminant analysis: determination of a function which separates two sets of curves
- Functional linear models:
  - The response variable $x(t)$ is a function
  - At least one explanatory variable is a function
  - both the above
- Functional analysis of variance
- Linear differential equations: principal differential analysis

Durham University

Summary

- Data treated as functions, with functional summaries and functional analogues to classical methods.
- Methods extend to more complicated objects such as shapes in many dimensions.
- Tend to work with smoothed versions of the data and not the raw data: understates actual variability
- Powerful methods with much elegant theory
- Needs some experience to make interpretations, may be a danger in post-hoc rationalization
- Useful exploratory tool. If the voice of the data is strong, can also be a useful inferential tool.
- Used R and fda package for computation and plots

Durham
University

References and resources

- J.O. Ramsay & B.W. Silverman (2002) Applied functional data analysis: methods and case studies. New York: Springer. Contains a lot of examples, but little theory.
- http://www.stats.ox.ac.uk/ silverma/fdacasebook/ Data and some (Matlab, R, S+) code for the case studies book.
- J.O. Ramsay & B.W. Silverman (2005) Functional data analysis. New York: Springer. Second edition, much expanded
- http://ego.psych.mcgill.ca/misc/fda/ Main FDA website, containing R library package and data sets.
- J.O. Ramsay & B.W. Silverman (2009) Functional data analysis with R and Matlab. New York: Springer.