Functional principal components analysis

1. The goal of principal components analysis

- PCA is usually used when we want to find the dominant *modes of variation* in the data, usually after subtracting the mean from each observation.
- We want to know how many of these modes of variation are required to achieve a satisfactory approximation to the original data.
- It may be assumed that keeping only dominant modes will improve the signal–to–noise ratio of what we keep.
- We usually want to know what these modes represent in terms that we can explain to non–statisticians. Rotation of the principal components can help at this point.

2. Defining functional PCA

- Let's see what changes when we go from the multivariate version to the functional version.
- The short answer: Summations change into integrations

Multivariate PCA

1. Find *principal component weight* vector $\boldsymbol{\xi}_1 = (\xi_{11}, \ldots, \xi_{p1})'$ for which the *principal components scores*

$$f_{i1} = \sum_j \xi_{j1} x_{ij} = \boldsymbol{\xi}_1' \mathbf{x}_i$$

   maximize $\sum_i f_{i1}^2$ subject to

$$\sum_j \xi_{j1}^2 = \|\boldsymbol{\xi}_1\|^2 = 1.$$

2. Next, compute weight vector $\boldsymbol{\xi}_2$ with components ξ_{j2} and principal component scores maximizing $\sum_i f_{i2}^2$, subject to the constraint $\|\boldsymbol{\xi}_2\|^2 = 1$ and the additional constraint

$$\sum_j \xi_{j2}\xi_{j1} = \boldsymbol{\xi}_2'\boldsymbol{\xi}_1 = 0.$$

3. and so on as required.

Functional PCA

1. Find *principal component weight* function $\xi_1(s)$ for which the *principal components scores*

$$f_{i1} = \int \xi_1(s) x_i(s)\, ds$$

maximize $\sum_i f_{i1}^2$ subject to

$$\int \xi_1^2(s)\, ds = \|\boldsymbol{\xi}_1\|^2 = 1.$$

2. Next, compute weight function $\xi_2(s)$ and principal component scores maximizing $\sum_i f_{i2}^2$, subject to the constraint $\|\boldsymbol{\xi}_2\|^2 = 1$ and the additional constraint

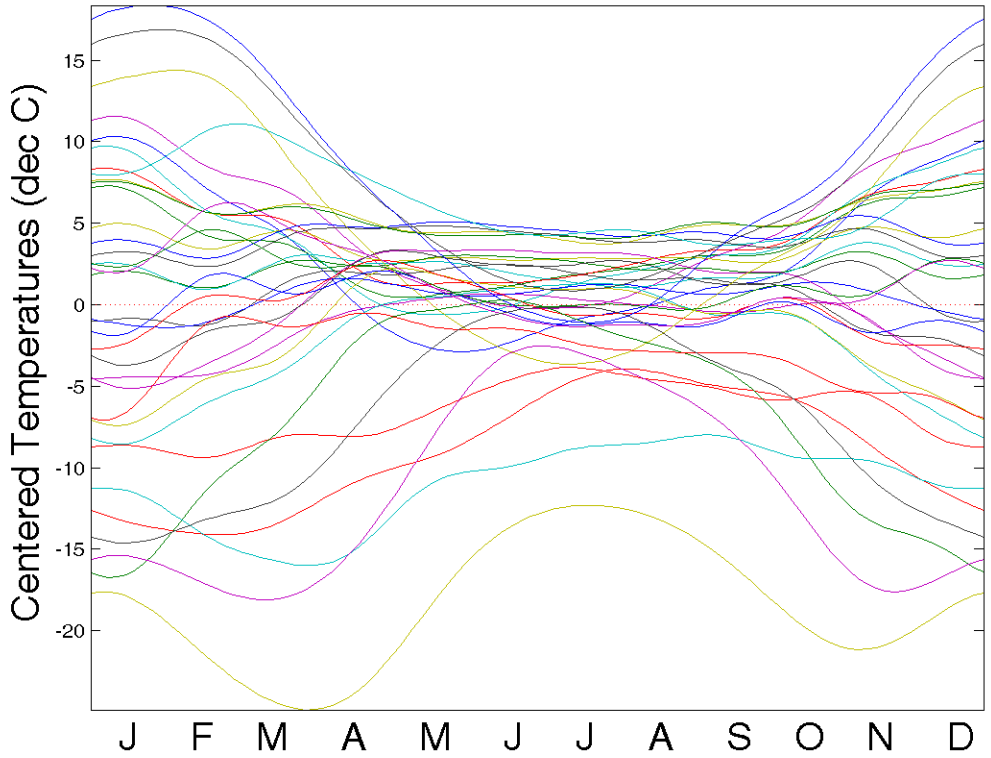
$$\int \xi_2(s)\xi_1(s)\, ds = 0.$$

3. and so on as required.

3. A PCA of monthly temperature curves

- We have 30-year average temperatures for each month and for each of 35 Canadian weather stations.

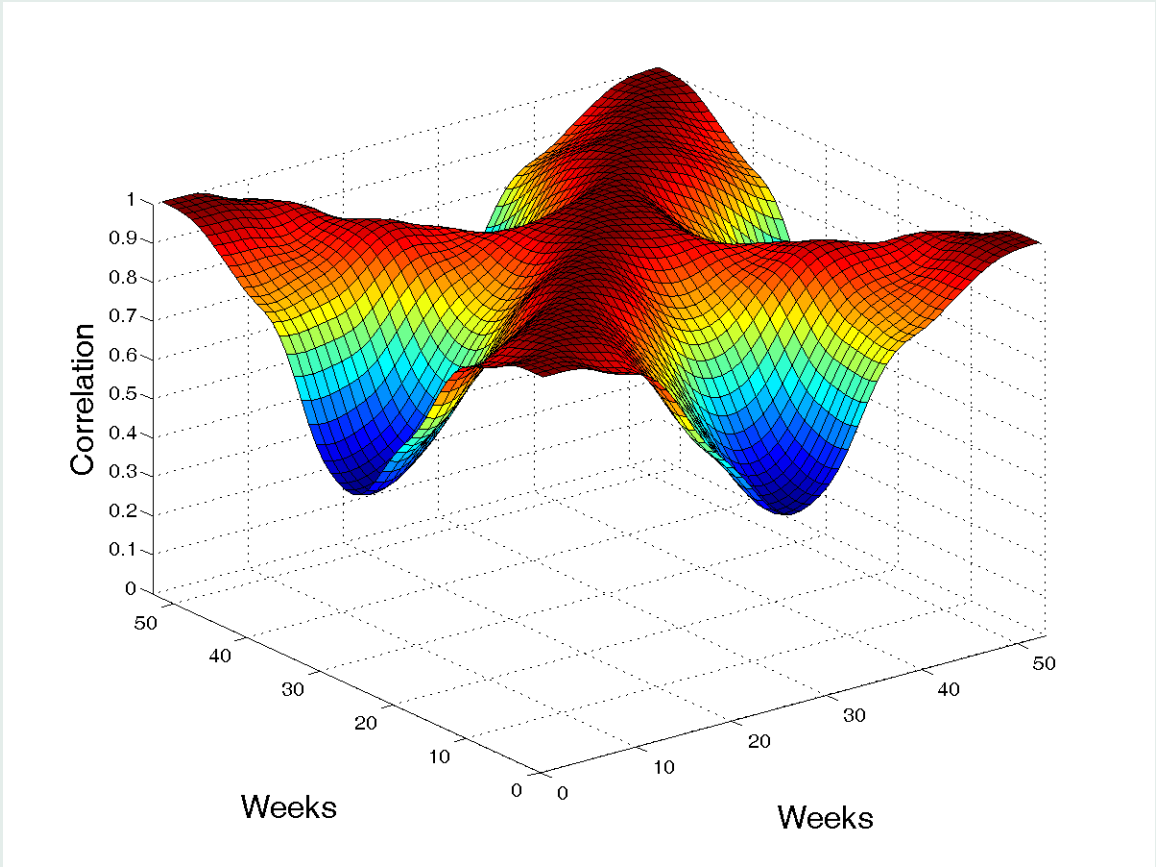
The centered monthly temperature curves



What do we see?

- An impression that some curves are high (warm) and that some curves are low (cold).
- Also that some curves have larger variation between summer and winter than others.
- How much of the variation do these two types of variation account for?

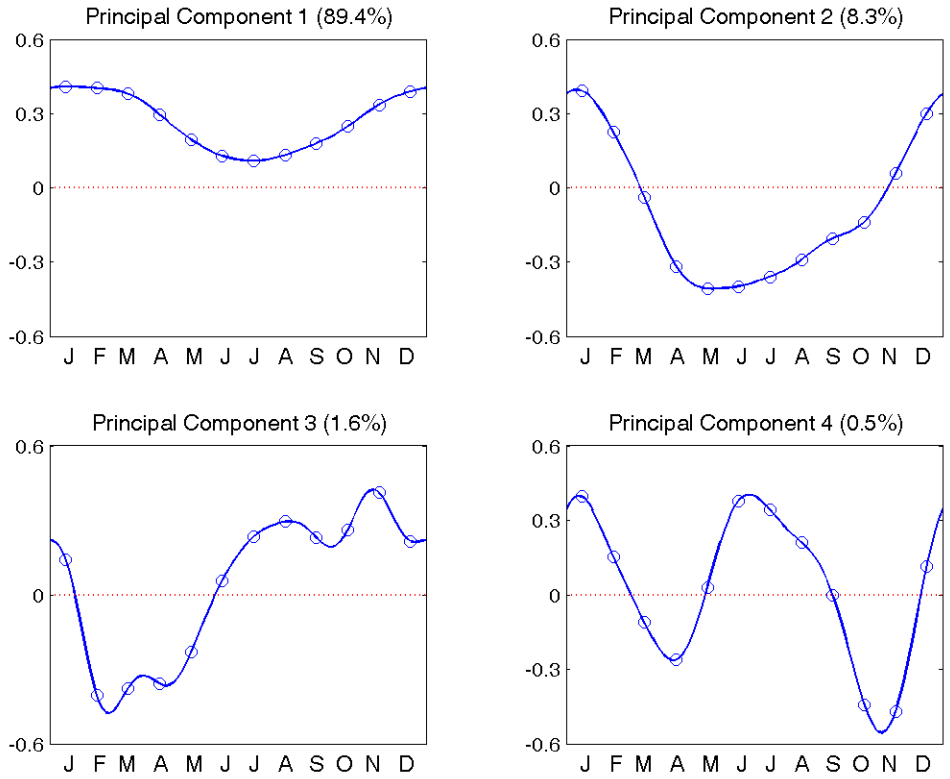
The correlation surface



What do we see?

- The diagonal ridge corresponding to unit correlation between temperatures at identical times.
- The ridge perpendicular to this corresponding to correlations between temperatures symmetrically placed around mid–summer.
- Correlations fall off much more rapidly for times symmetric about March and September 21.

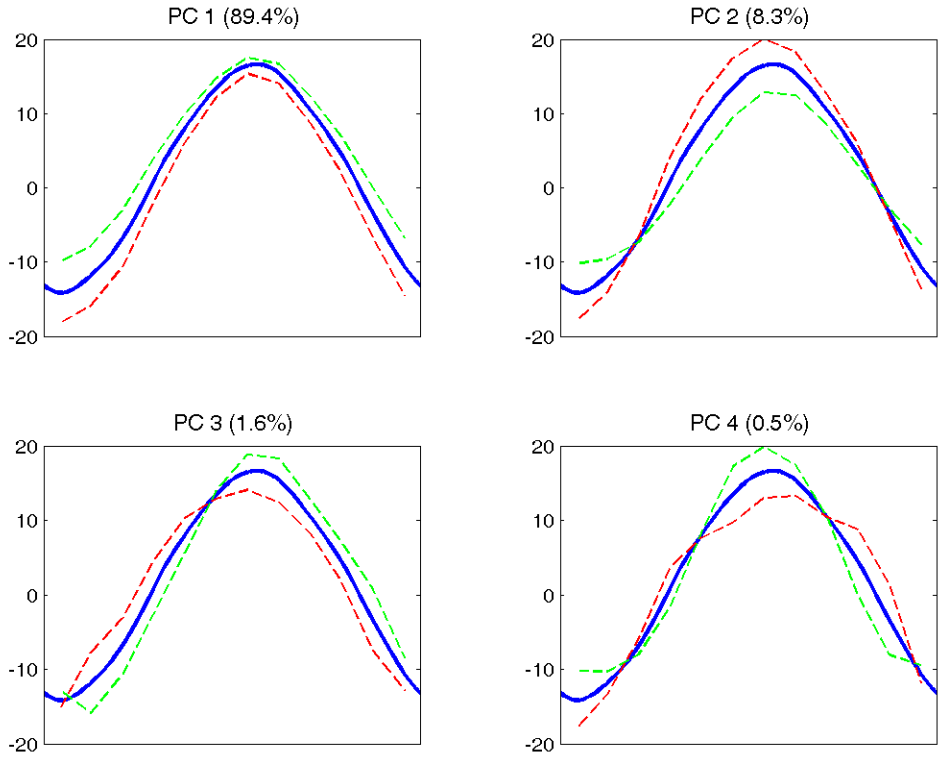
The first four principal components



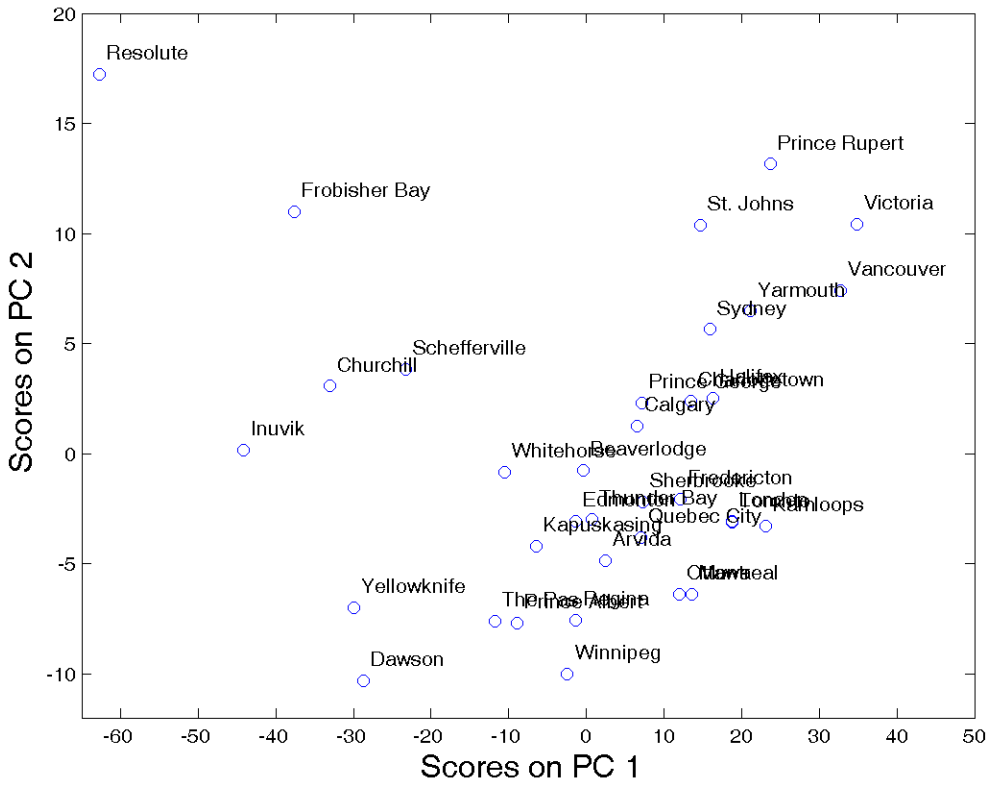
What do we see?

- The two components that we saw in the centered curves account for about 98% of the variation.
- The first four components account for 99.8% of the variation.
- The first four components tend to look like linear, quadratic, cubic and quartic polynomials, respectively. Why is that?
- It can help to plot the components by adding and subtracting a multiple of them from the mean function.

The first four principal components +/- mean



The first two principal component scores



What do we see?

- Most stations are along a curved line running from lower center to top right.
- At the top end of the banana are maritime stations with less variation between winter and summer, and high average temperatures.
- At the lower end are the continental stations with large seasonal variation and lower average temperatures.
- The Arctic stations are in their own space with large seasonal variation and very low average temperatures.

4. Perspectives and rotations

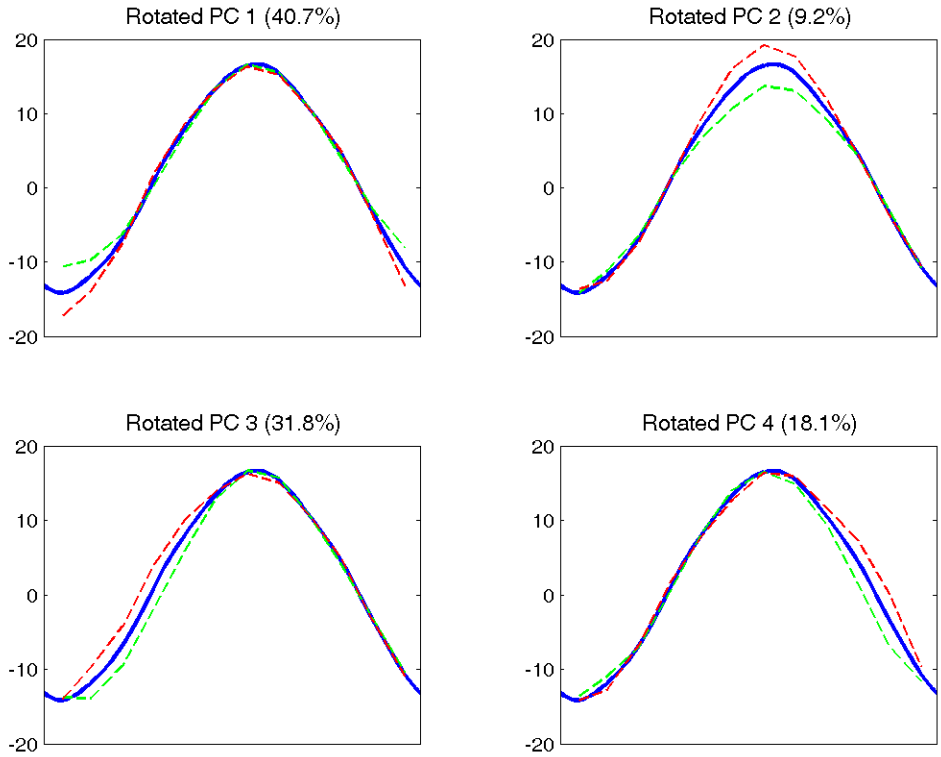
Principal components as empirical orthogonal functions

- We can think of principal components as a set of orthogonal basis functions constructed so as to account for as much variation at each stage as possible.
- In fact, they are often used as just that: A compact basis for approximating the data with as few basis functions as possible.
- They come out looking like polynomials of increasing degree because dominant variation tends to be smooth (i. e. nearly constant or linear), and subsequent components pick up variation that declines in smoothness, and is also required to be orthogonal to previous components. Just like orthogonal polynomials!

Rotating principal components

- Once we have a set of orthogonal components spanning as much variation as we desire, we can always rotate these orthogonally to get a new set spanning the same space.
- The advantage is that rotated components may be easier to interpret.
- The VARIMAX rotation method is often used in the social sciences to improve interpretability.
- Functional principal components can be rotated in this way as well.

Rotated principal components for temperature



What do we see?

- The total variation accounted for remains the same, 99.8%.
- The first two components now account for a less overwhelming amount of the variation.
- Each rotated component now accounts for departure from the mean for a small part of the year.
- These are much easier to interpret. Components 1 and 3 are the most important, and account for deviation from the mean in mid–winter and in the fall, respectively.

How many principal components can be computed?

- In the multivariate case, the upper limit is the number of variables.
- In the function case, “variables” correspond to values of t, and there is no limit to these.
- Instead, the upper limit is the number N of observations, or $N - 1$ if the functions are centered.
- But in some cases, the number of basis functions K will be less than N, and in this case K is the upper limit.
- We usually stop far short of either of these limits, however.

What if the functions are themselves multivariate?

- This often arises if the functions are spatial coordinates, $[X(t), Y(t), Z(t)]$ or angular coordinates. Then we want to study their simultaneous variation, rather than separately.
- The solution is simple: Make a single synthetic function by joining them together, compute it's principal components, and separate out the parts belong to each coordinate.

What if I had a mixture of functional and scalar variables?

- This often happens. We could study the components of simultaneous variation in temperature profiles and log total annual precipitation, for example.
- Or the simultaneous variation in growth acceleration curves and the parents' adult stature.
- Ramsay and Silverman (1997, 2004) show that this, too, can be converted to a matrix eigenequation.

5. How are functional principal components computed?

- In multivariate statistics, we solve the eigenequation

$$\mathbf{V}\boldsymbol{\xi} = \rho\boldsymbol{\xi}$$

  where

  – $\mathbf{V}$ is the sample variance-covariance matrix

$$\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$$

    where, in turn, $\mathbf{X}$ is the centered data matrix.

  – $\boldsymbol{\xi}$ is an eigenvector of $\mathbf{V}$.

  – ρ is an eigenvalue of $\mathbf{V}$.

- Usually, however, we actually use the correlation matrix $\mathbf{R}$ instead of $\mathbf{V}$ so as to eliminate uninteresting scale differences between variables.

What is the function version of the eigenequation?

- Let
$$v(s,t) = N^{-1} \sum_{i}^{N} x_i(s) x_i(t)$$
where usually functions $x_i(t)$ have been first centered.
- $v(s,t)$ is the sample *variance-covariance function*.
- The functional eigenequation is
$$\int v(s,t)\xi(t)\,dt = \rho\xi(s)$$
- ρ is still an eigenvalue, but now $\xi(s)$ is an *eigenfunction* of the variance-covariance function.
- There is much less reason for using the correlation function $r(s,t)$ since function values all have the same units or scale.

How do we solve for pairs of eigenvalues and eigenfunctions?

- Suppose that the observed functions are expanded in terms of a vector $\boldsymbol{\phi}(t)$ of K basis functions

$$\mathbf{x}(t) = \mathbf{C}\boldsymbol{\phi}(t)$$

- and the jth eigenfunction the expansion

$$\xi_j(s) = \mathbf{b}_j'\boldsymbol{\phi}(s) \ .$$

- Substituting these expansions into the equation for $v(s,t)$ gives us

$$v(s,t) = N^{-1}\boldsymbol{\phi}'(s)\mathbf{C}'\mathbf{C}\boldsymbol{\phi}(t)$$

- The eigenequation becomes

$$N^{-1}\boldsymbol{\phi}'(s)\mathbf{C}'\mathbf{C}\int\boldsymbol{\phi}(t)\boldsymbol{\phi}'(t)\,dt\,\mathbf{b}_j=\rho\boldsymbol{\phi}'(s)\mathbf{b}_j$$

- Define order K matrix

$$\mathbf{J}=\int\boldsymbol{\phi}(t)\boldsymbol{\phi}'(t)\,dt$$

so that the eigenequation is now

$$N^{-1}\boldsymbol{\phi}'(s)\mathbf{C}'\mathbf{C}\mathbf{J}\mathbf{b}_j=\rho\boldsymbol{\phi}'(s)\mathbf{b}_j$$

- This equation has to be true for all argument values s, and consequently,

$$N^{-1}\mathbf{C}'\mathbf{C}\mathbf{J}\mathbf{b}_j=\rho\mathbf{b}_j$$

- subject to the constraint $\|\xi\|^2=1$, which becomes

$$\mathbf{b}_j'\mathbf{J}\mathbf{b}_j=1\ .$$

- if we define

$$\mathbf{u}_j = \mathbf{J}^{1/2}\mathbf{b}_j$$

- then we have the symmetric eigenequation

$$N^{-1}\mathbf{J}^{1/2}\mathbf{C}'\mathbf{C}\mathbf{J}^{1/2}\mathbf{u}_j = \rho\mathbf{u}_j$$

  subject to the constraint

$$\mathbf{u}_j'\mathbf{u}_j = 1$$

- We can then use standard software to solve for the eigenvectors $\mathbf{u}_j$ and back–solve to get the required coefficient vectors

$$\mathbf{b}_j = \mathbf{J}^{-1/2}\mathbf{u}_j$$

  for computing the eigenfunctions $\xi_j(s)$.

Suppose that I wanted to impose a roughness penalty on $\xi_j(s)$

- Indeed. If the data are rough, the eigenfunctions will be, too, unless we force them to be smooth.
- Skipping some technicalities, if we penalize $\|D^2\xi\|^2$, for example, we find that ξ satisfies the modified eigenequation

$$\int v(s,t)\xi(t)\,dt = \rho[\xi(s) + \lambda D^4\xi(s)]$$

- This, too, can be converted to an equivalent matrix equation that is solvable with standard software.