A stroll along the beach

Charlotte Wickham
University of California, Berkeley

The data

The data consist of a time series of images of Pauanui Beach on the Coromandel Peninsula in New Zealand.

For this analysis I am using:
- midday images over two years (Jan 2004 to Dec 2005)
- transformed to be greyscale
- rescaled to 190 x 143 pixels (27170 pixels in total).

We can think of each image as a matrix of 190 x 143 numbers (pixels) each having a value between 0 and 1 (black to white).

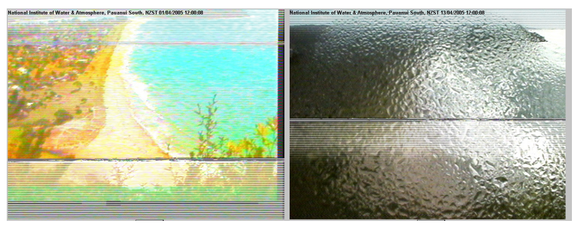
The images are collected by the National Institute for Water and Atmospheric Research (NIWA) and I thank them for supplying them to me.

Data cleaning

Problems with the image series were diagnosed by looking at time series of the mean intensity and the variance of the intensity within each image.

Images after the 1st of April had unusually high day to day variation in intensity. Some examples of these images are to the right. It is clear the camera was malfunctioning. These images had to be dropped from further analysis.

Days with very low mean intensity were blank images and were removed from the series. Days with unusually low variance in intensity and a moderately low mean tend to be stormy days. These were kept in the series.

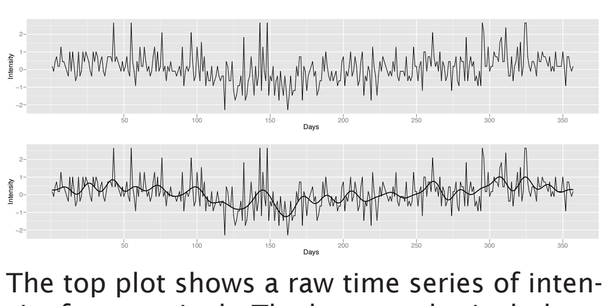


Examples of images after April 1st 2005. The camera is malfunctioning.

Can functional data analysis give us insight into this data?

Functional principal components analysis

Each pixel is treated as an individual time series of intensity. We assume that the series we see is a noisy version of some underlying smooth function. We first smooth each pixel's time series using a Fourier basis of size 51. This serves two purposes. Firstly, to approximate the smooth function and secondly to reduce the dimensionality of the problem.



The top plot shows a raw time series of intensity for one pixel. The bottom plot includes a smoothed version of the series.

We then perform functional principal components analysis. This is essentially the same as principal components but instead of searching in a vector space we are now in a function space.

In practice the two are identical if we treat the pixels as observations and the coefficients on the Fourier basis as the variables.

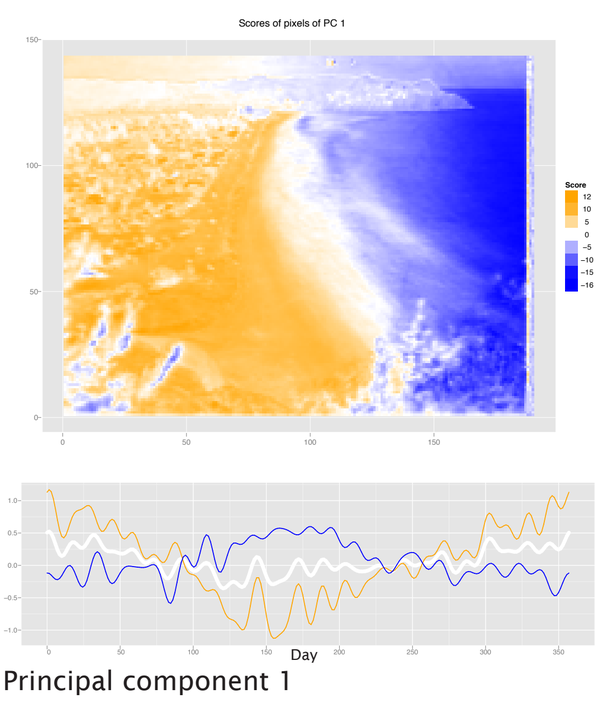
We hope that the principal components give us insight into the patterns of variation amongst the pixels over time.

Results

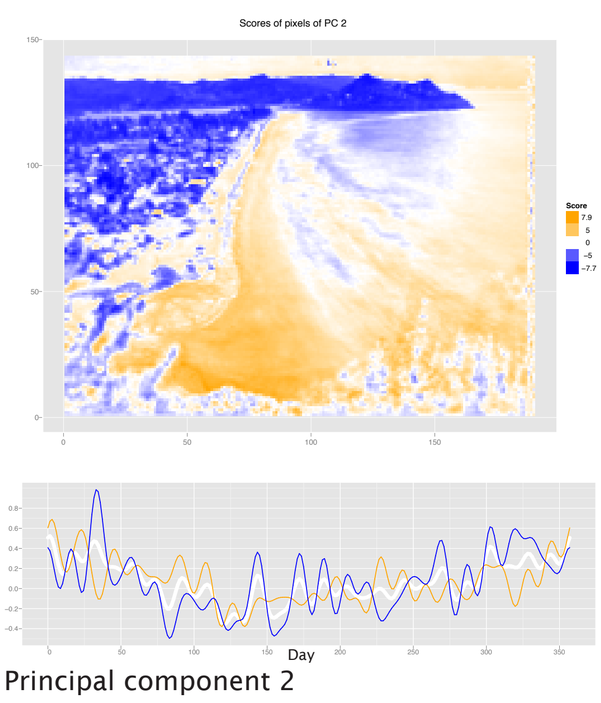
The images below show the results of applying FPCA to the 27170 pixels. The top image shows each pixel's score on the component. The bottom plot illustrates the functional form of the component.

The white line is the average across all pixels. The orange line is the result of adding a multiple of the component to the average.

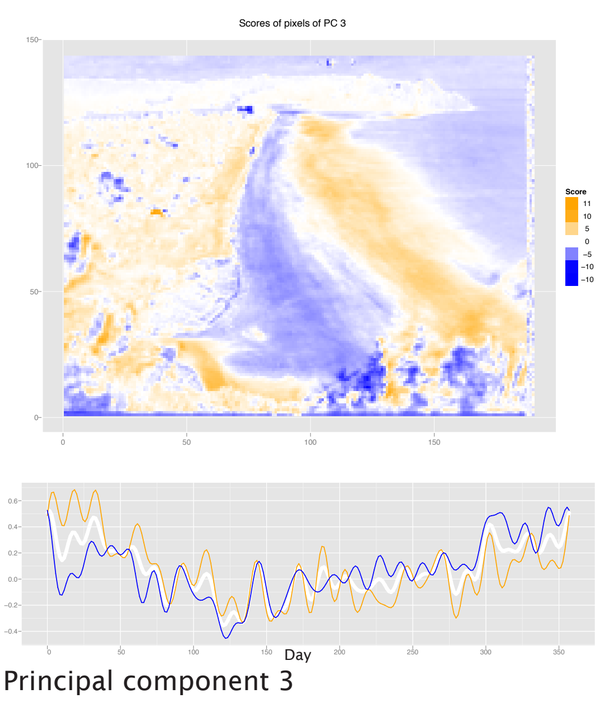
The blue line is the result of subtracting a multiple of the component.



Principal component 1



Principal component 2



Principal component 3

Principal component 1
Divides the image down the middle. Pixels that score low on this component are lighter in the winter and correspond to the water.

Principal component 2
The scores still divide the image into physically interpretable areas. The functional form of the component is harder to interpret.

Principal component 3
Pixels that score highly on this component are lighter than average in the first half of the year and dimmer in the second.

Summary

FPCA does give insight into image series.

Each pixel is treated independently yet when looking at the scores we pick out actual physically interpretable areas.

References

All analyses were performed in R using the packages pixmap, fda and ggplot.

Roger Bivand, Friedrich Leisch, and Martin Machler. *pixmap: Bitmap Images ("Pixel Maps")*, 2006. R package version 0.4-5.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

J. Ramsay and B. Silverman. *Applied Functional Data Analysis*. Springer-Verlag, 1997.

J. O. Ramsay and Hadley Wickham. *fda: Functional Data Analysis*, 2006.

Hadley Wickham. *ggplot: An implementation of the Grammar of Graphics in R*, 2006.

FPCA reveals patterns of variation among the pixels.