

HW06 FDA spatial PC

Drew Remmenga

2022-10-30

- You are encouraged to use web resources and class materials
- You can work together but hand in your own work.
- Be spare in what you include and you will lose credit if you include too much extraneous output or information. All parts questions count for an equal number of points.
- Any subproblems marked GRAD are required for 500 level students but will serve as an extra credit question for the 400 level students.
- Please send me email if you have questions or any concerns. nychka@mines.edu
- Hand in your work in pdf format in Gradescope. You can keep the questions as part of what you hand in but you should begin your *answer* on a separate page. You can use `\newpage` to create a page break in your work.

Reformatting text and code that are not your answers

It is harder to grade homework when all of the introduction and question text are included. Either comment out the answers without just deleting them or put your answers in a different color.

To use the html commenting format:

```
<!--  
This text is now commented out and will not be part of the rendered output.  
-->
```

To change the color you need to have Latex working:

```
\textcolor{magenta}{This is a magenta block of text. }
```

And the rendered version in pdf:

This is a magenta block of text.

Points

All subsections of the problems count equally for 10 points:

- 400 level
- 500 level

Some setup

Change the directory path below to be for your PC/Mac

```
suppressMessages(library( fields))
```

```
## Warning: package 'fields' was built under R version 4.1.3
```

```
## Warning: package 'spam' was built under R version 4.1.3
```

Problem 1

Load the file `HW06StationComplete.rda` , a subset of the 2021 Colorado data set that includes the locations, elevations and times.

These data have been reduced and filled so that there are complete observations for every hour and every day that is reported. **stationComplete** is a 3 dimensional array with indices, hour, day, and station.

The day of the year for these complete records is in the vector **tday**. Note that there are only 187 of these days available out the possible 365 and is due to the cases where the number of missing was not feasible to fill in.

s are the station locations in longitude/latitude.

elev are the station elevations in meters.

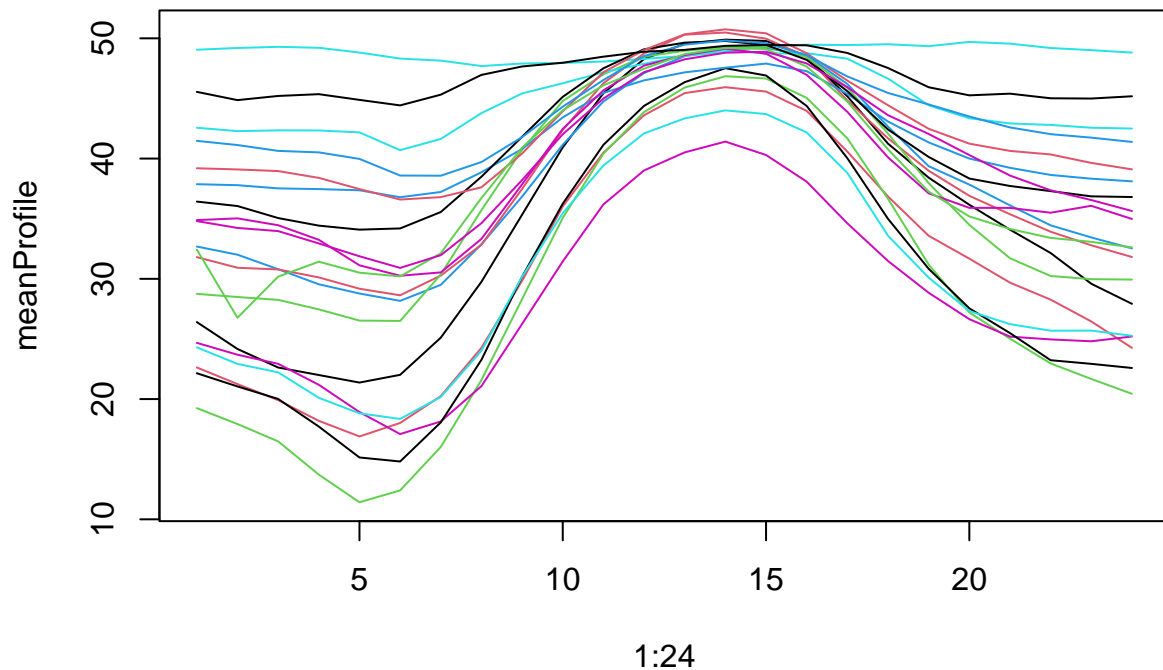
The code below loads the data and finds the mean daily profile for each station. Note the use of the `apply` function here and how it is set to average over the middle dimension, days.

```
load("HW06StationComplete.rda")
dim( stationComplete)
```

```
## [1] 24 187 19
```

```
# plot of mean day cycle over the year for all stations
```

```
meanProfile<- apply( stationComplete, c(1,3), mean)
matplot( 1:24, meanProfile, type="l", lty=1 )
```



1(a)

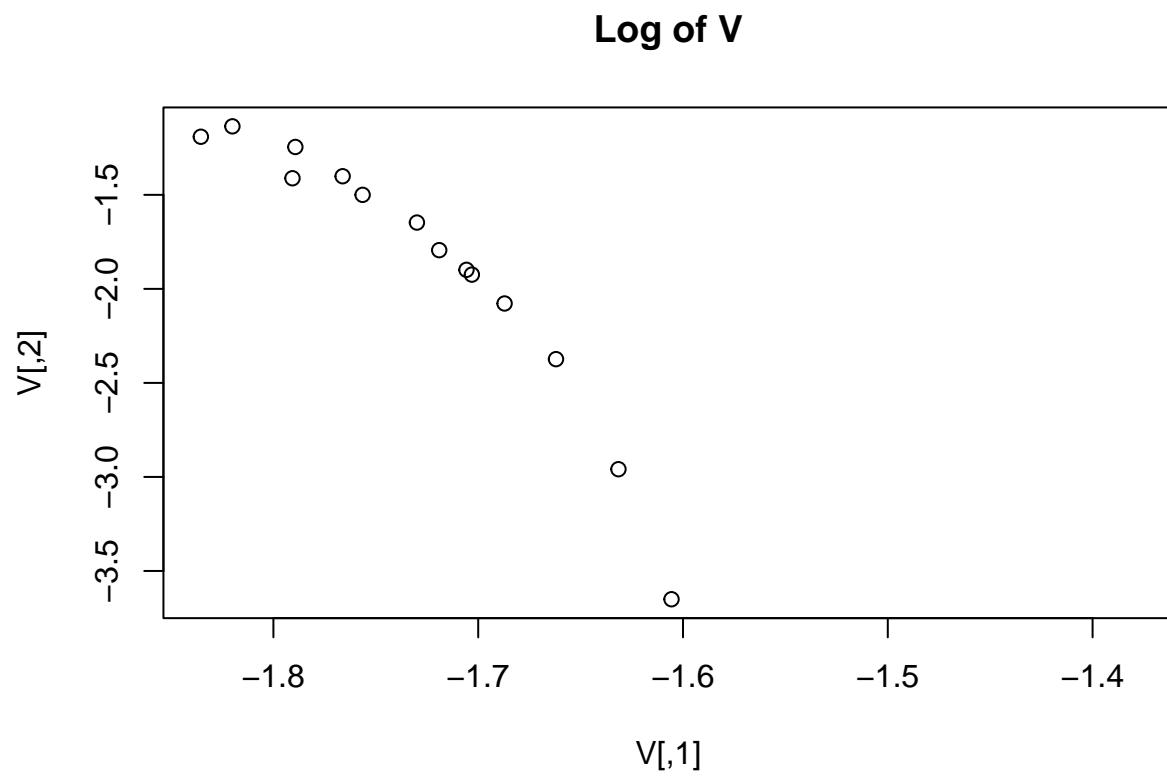
Compute an SVD on $\mathbf{t}(\text{meanProfile})$, UDV^T So that the rows of U index stations and the columns of V are the normalized basis functions. In this case to make the interpretation easier it makes sense to multiply “ U ” and “ V ” by -1 from the default version provided by `svd`.

Plot the log singular values and comment on if there is sharp drop at any points aka as a “knee”.

```
svd=svd(t(meanProfile))
v=svd$v*-1
V=log(v)
```

```
## Warning in log(v): NaNs produced
```

```
plot(V,main="Log of V")
```

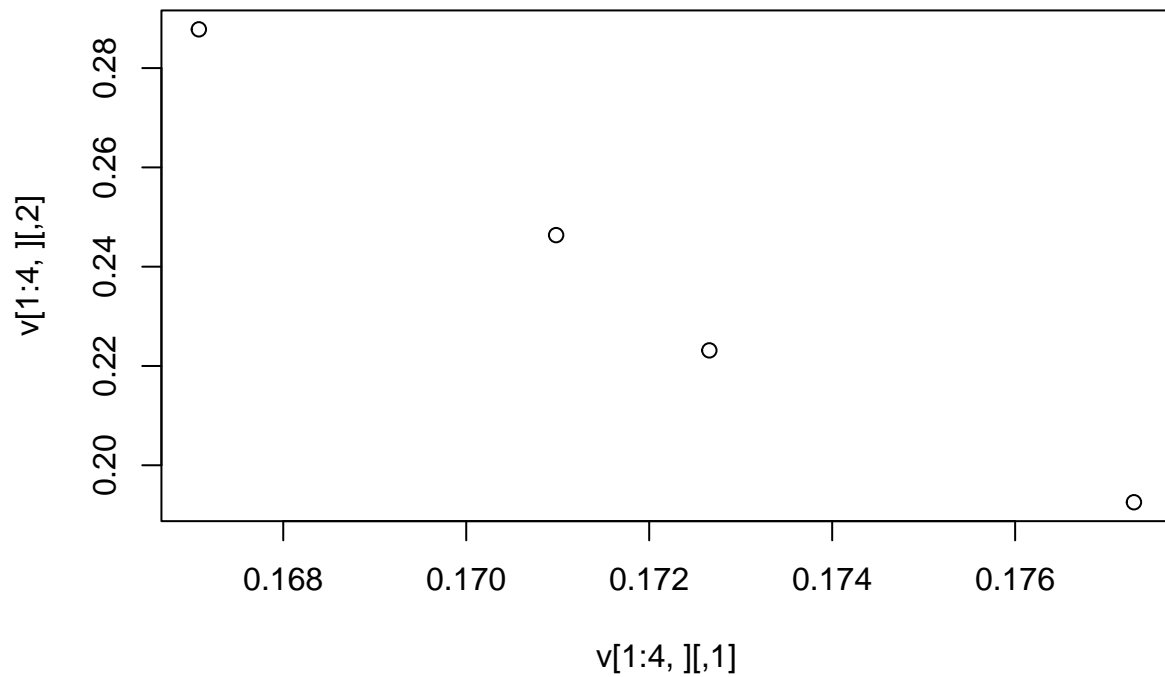


There is a knee at the fourth V value.

1(b)

- Plot the first 4 basis functions from this decomposition in the normalized form from the “V” matrix

```
plot(v[1:4,])
```

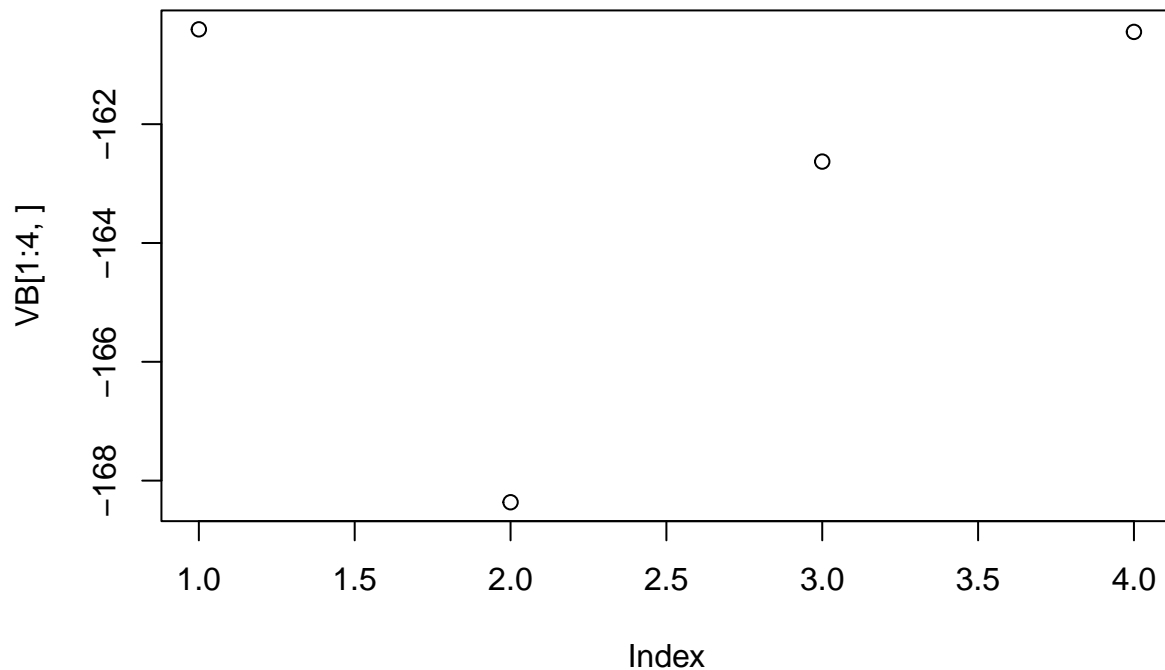


- Explain why in this normalized form the squared values of each basis function sum to 1. That is `sum(V[,k]^2)` is one for all `k`. The matrices are orthonormal.

1(c)

- Weight the basis functions by their singular values as `VB <- V*%D` and plot the first 4 of them.

```
VB <- svd$v*%svd$d
plot(VB[1:4,])
```



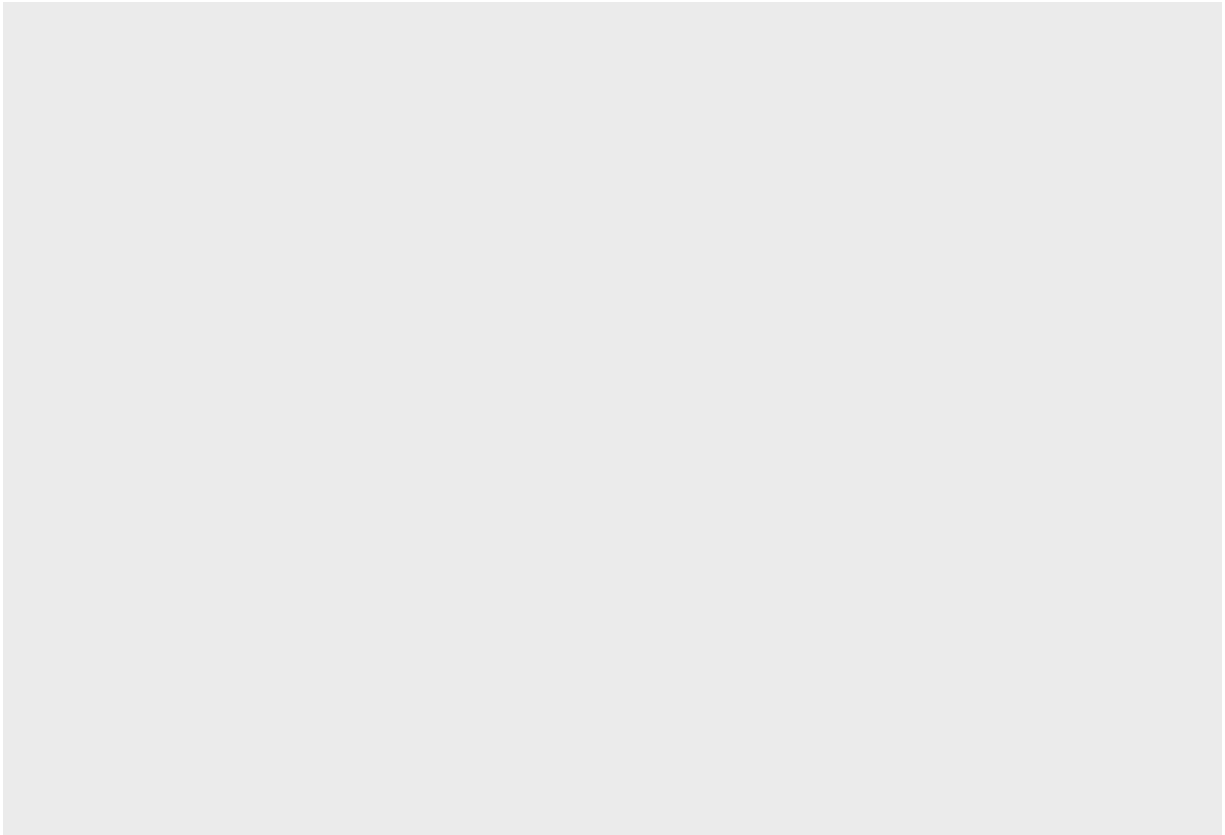
- Comment on how the shape of the second basis function can modify the first. The second basis function covers so much of the data that the first doesn't need much modification.

Hint: To see this make a plot of three curves the first basis function, the first minus B times the second and the first plus B times the second. I call these a *plus/minus plot* and for these data I found a value of value $B=1.0$ makes sense based on the size of the U coefficients.

1(d)

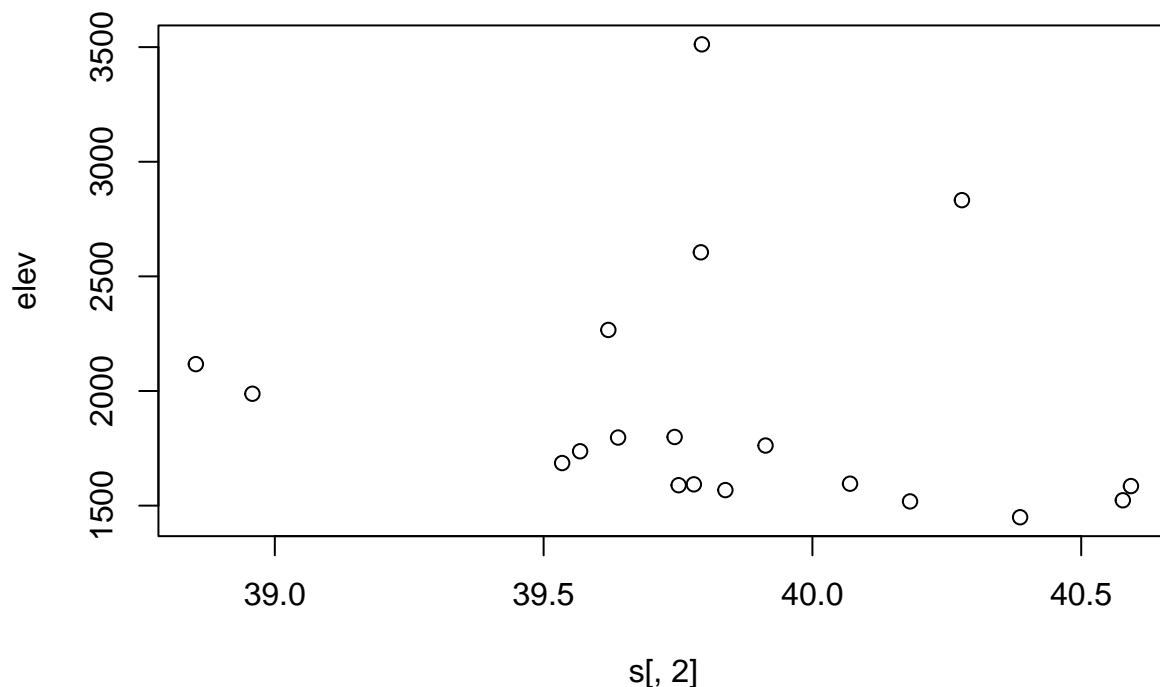
- Make a `bubblePlot` of the second coefficients over space. Does there seem to be some spatial pattern? Choose the colors and bubble size to make a good graphic.

```
library(ggplot2)
ggplot(s)
```



- Plot the second coefficient against elevation (the **elev** vector) and comment on any patterns.

```
plot(s[,2],elev)
```



PROBLEM 2

Use the code below to create

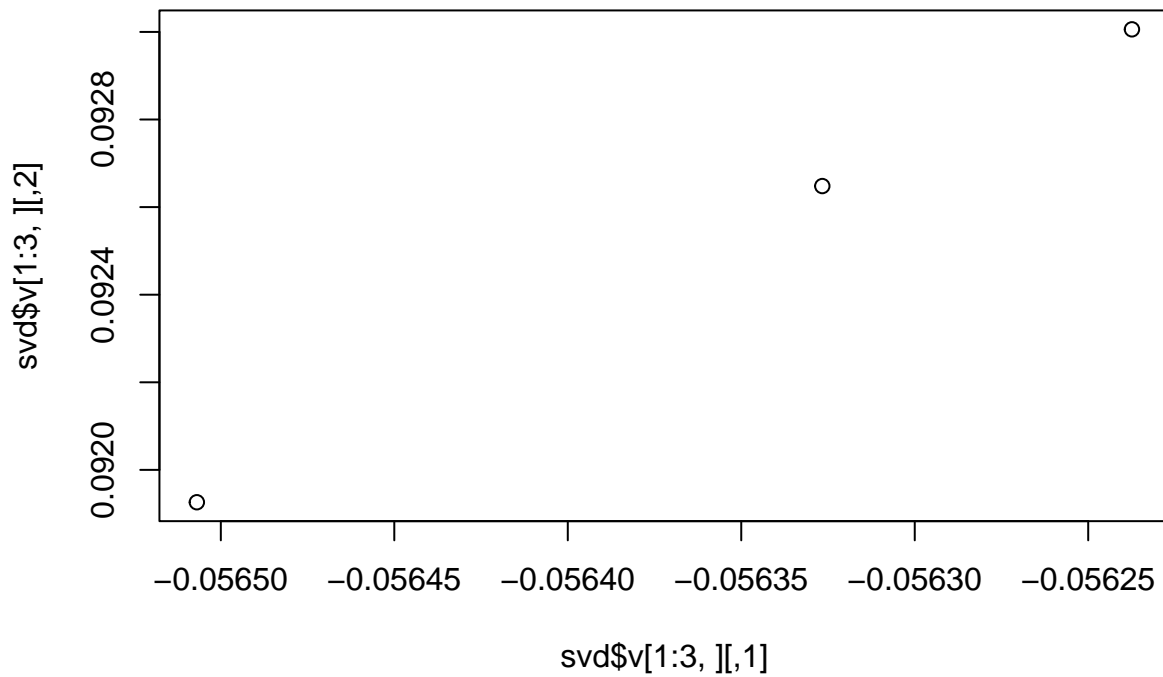
a data set, `dayProfile0` for the max ozone for each day and for each station. (Rows index day and columns index station.) Also create `dayProfile` by smoothing each of the 19 time series using Tps with 10 effective degrees of freedom in the smoothed curves.

```
dayProfile0<- apply( stationComplete, c(2,3), max)
dayProfile<- matrix( NA, nrow(dayProfile0), 19)
for( k in 1:19){
  fit<- Tps( tday, dayProfile0[,k], df = 10)
  dayProfile[,k]<- fit$fitted.values
}
```

2 (a)

- Find the SVD for `t(dayProfile)`, plot the first three basis functions weighted by their singular values and comment on their shapes. Be sure to use `tday` for the time on the horizontal axis and I suggest multiplying U and V by -1.

```
svd=svd(t(dayProfile))
plot(svd$v[1:3,])
```

- Similar to 1(c) make a plus/minus plot for the first and second weighted basis functions. Again, $B=1.0$ is a good choice to make the plot.

2 (b)

How does the shape of these time series depend on the elevation of the station? They are directly related.

2(c) GRAD

Plot together and on a log scale the singular values for `dayProfile0` with `dayProfile`. Explain why the smoothed data may have singular values that decrease more quickly.