CHAPTER 13

COMPUTATIONAL EXERCISES ON FIELD SIGNIFICANCE

In this homework set, you will perform a series of field significance tests. The first step of this homework is to obtain the following files from the class website. It is good practice to put all of your data in a single directory (say "data"), and all your auxiliary programs into a single directory that can be referred to in future R programs.

`Rcode.exercise.Chapter13.R`	main program: reads data and computes EOFs
`cca.data.clim763.DJF.RData`	data (will be used in future homeworks)
`eof.latlon.R`	auxiliary file: calculate EOFs
`plot_latlon_v4.R`	auxiliary file: plot spatial maps
`gev.R`	auxiliary file: function to solve eigenvalue problem
`index.climate.v2.R`	auxiliary file: function to compute NINO3.4 index

Your starting point is the R program `Rcode.exercise.Chapter13.R`, which does the following things things:

- It reads the data file `cca.data.clim763.DJF.RData` and loads them in the arrays `tem.data` and `sst.data`. Inside `Rcode.exercise.Chapter13.R`, you need to modify the variable `dir.data` to correspond to the directory to which you downloaded the data.

- It reads the auxiliary programs (e.g., `eof.latlon.R` and `plot_latlon_v4.R`). You will need to modify `dir.Rlib` to correspond to the directory to which you downloaded the auxiliary programs.
- It computes the nino3.4 index from `sst.data`.
- It also computes the EOFs of U.S. temperature using `eof.latlon`, which creates the list variable `tem.eof`. The EOFs are `tem.eof$eof[space,neof]`, and the pc time series are `tem.eof$pc[time,neof]`.
- It plots the leading EOF and PC of North American Temperature and SST. The resulting figure should look like fig. 13.1.

Note: temperature over the ocean is set to `NA`. Also, temperature is read as an array with dimension `nlon, nlat, time`, but eventually is "reshaped" into an array with dimension `nlon*nlat, nyrs`. If you do not understand what this means, please see me!

The purpose of this homework is to test the hypothesis that tropical Pacific SSTs influence North American temperature in DJF. You will begin by looking at the correlation map between temperature and NINO3.4. This is very convenient to plot in R. First, we compute the correlation between NINO3.4 and every grid point of temperature using the command

```
cor.map = cor(t(tem.data),nino34)
```

The R-function `cor` assumes data matrices are in the form TIME x SPACE, so we must take the transpose of `tem.data` (which is an array with dimensions SPACE x TIME) to use `cor`. To plot this, we then use the command

```
plot_latlon_v4(lon,lat,cor.map).
```

To "maskout" insignificant correlations, we use the commands

```
cor.crit = 2/sqrt(nyrs)
cor.map[abs(cor.map)<cor.crit] = NA
plot_latlon_v4(lon,lat,cor.map)
```

The resulting plot should look like fig. 13.1 in the notes.

Assignment

Exercise 13.1. Write an R code that applies the Livezey-Chen field significance test. The preamble of the function should be the following:

```
livezey.chen = function(lon,lat,xdata,ydata,ntrials=1000,alpha=0.05) {
### PERFORMS THE LIVEZEY-CHEN FIELD SIGNIFICANCE TEST
# INPUT:
#   LON[NLON]: LONGITUDE OF THE FIELD DATA
#   LAT[NLAT]: LATITUDE OF THE FIELD DATA
#   XDATA[NLON,NLAT,NTOT]: DATA ARRAY FOR THE FIELD
#   YDATA[NTOT]: REFERENCE TIME SERIES
#   NTRIALS: NUMBER OF MONTE CARLO TRIALS (DEFAULT = 10000)
#   ALPHA: SIGNIFICANCE LEVEL OF THE TEST (DEFAULT = 5%)
# OUTPUT:
#   HISTOGRAM OF AREAS WITH SIGNIFICANT CORRELATIONS
#   LIST:
#     $CRIT.NULL: SIGNIFICANCE THRESHOLD FROM LIVEZEY-CHEN TEST
#     $SIG.AREA.OBS: PERCENTAGE AREA WITH SIGNIFICANT CORRELATIONS
```

Apply the Livezey-Chen field significance test to test whether NINO3.4 is related to DJF temperature over North America. Make a histogram of the Monte Carlo results.

Be sure to express the answer in terms of percentage area with significant correlations. Suggestion: create another field that gives the fractional area at each grid point, then sum over those points. For instance:

```
area.frac    = rep(cos(lat*pi/180),each=nlon)
lgood        = !is.na(xdata[,1])
area.frac[!lgood] = NA
area.frac    = area.frac/sum(area.frac,na.rm=TRUE)

cor.map      = as.numeric(cor(t(xdata),ydata))
sig.area.obs = sum(area.frac[abs(cor.map) > 2/sqrt(ntot)],na.rm=TRUE)*100
```

You should obtain a plot similar to fig. 13.4 of the lecture notes. Make a conclusion regarding whether the field is significant. □

Exercise 13.2. Write an R code that applies the permutation field significance test. The preamble of the function should be the following:

```
fieldsig.permutation = function(lon,lat,xdata,ydata,ntrials=1000,alpha=0.05) {
### PERFORMS PERMUTATION FIELD SIGNIFICANCE TEST
# INPUT:
#   LON[NLON]: LONGITUDE OF THE FIELD DATA
#   LAT[NLAT]: LATITUDE OF THE FIELD DATA
#   XDATA[NLON,NLAT,NTOT]: DATA ARRAY FOR THE FIELD
#   YDATA[NTOT]: REFERENCE TIME SERIES
#   NTRIALS: NUMBER OF TRIALS (DEFAULT = 10000)
#   ALPHA: SIGNIFICANCE LEVEL OF THE TEST (DEFAULT = 5%)
# OUTPUT:
#   HISTOGRAM OF AREAS WITH SIGNIFICANT CORRELATIONS
#   LIST:
#     $CRIT.NULL: SIGNIFICANCE THRESHOLD FROM LIVEZEY-CHEN TEST
#     $SIG.AREA.OBS: PERCENTAGE AREA WITH SIGNIFICANT CORRELATIONS
```

The `sample` function is very useful for generating a permutation. For instance, `sample(N)` generates a permutation of the integers 1-N. To ensure there are no accidental matches, the following loop may be helpful

```
temp  = TRUE
while(temp) {
        nperm = sample(nyrs)
        temp  = any(nperm == 1:nyrs)
}
```

Answer the same questions as the previous exercise. How does this result compare to that from the Livezey-Chen test? □

Exercise 13.3. Write a function that evaluates Mutual Information Criterion (MIC) for the field significance problem. MIC is a function of R-square, which can be computed from R functions as follows

```
r.square = summary(lm(y~x))$r.square
```

Show a plot of MIC versus the number of EOFs of U.S. temperature and identify the minimum. The figure should be similar to fig. 13.5 in the notes. □

Exercise 13.4. Write a code to evaluate the significance level for MIC. Include this in your function for MIC. Plot the significance level. □

Exercise 13.5. Write an R code that computes the False Discovery Rate. The preamble of this function should be the following:

```
fdr.fieldsig = function(y,x,fdr=0.1) {
## APPLIES FALSE DISCOVERY RATE TO A CORRELATION MAP
## INPUT:
#     Y[NTOT]: REFERENCE TIME SERIES
#     X[NTOT,MTOT]: FIELD TIME SERIES
#     FDR: FALSE DISCOVERY RATE
## OUTPUT:
#     LREJECT[MTOT]: LOGICAL VECTOR INDICATING REJECTIONS
```

Make a plot showing where the FDR implies rejection of the null hypothesis. The plot should be similar to fig. 12.7 of the notes. □

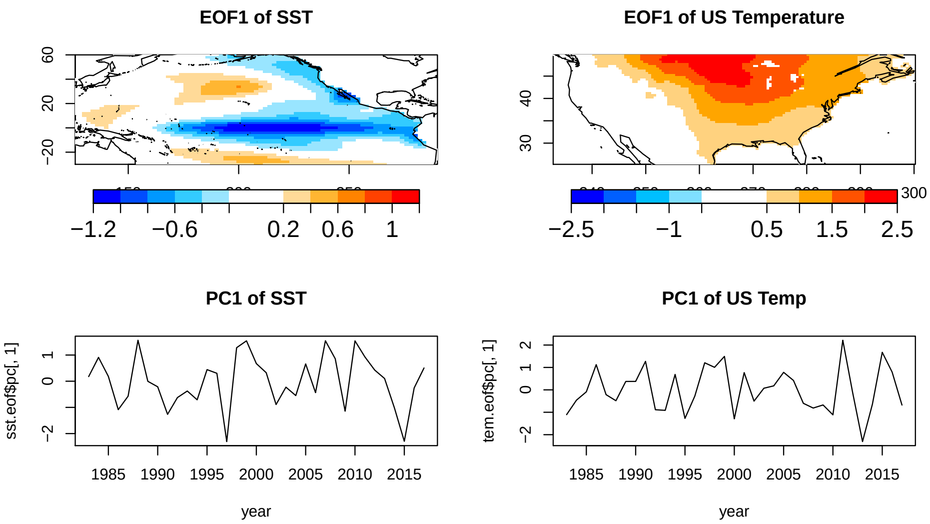

Figure 13.1 Leading EOF of DJF North American temperature during 1982-2017 (top) and correlation between DJF NINO34 index and DJF temperature during 1982-2017 (bottom).