CHAPTER 18

PREDICTABLE COMPONENT ANALYSIS

In this homework set you will apply Predictable Component Analysis (PrCA) to quantify the predictability of seasonal forecasts over North America from the CFSv2 Retrospective Hindcasts. The data set and codes for reading this data set are identical to those you used in the last homework for ANOVA.

To apply predictable component analysis, you must first compute the principal components of the data set. This can be done by pooling all ensemble members and years together as if it were one long time series. This will be done automatically by `eof.latlon`. Since the data set is dimensioned as

```
hcst.data[nlon,nlat,nens,nyrs]
```

On output, the PCs will have dimension `[nens*nyrs,neof]`. When desired, the PCs can be re-shaped into an array of dimension `[nens,nyrs,neof]`.

In this homework, you will eventually write an R function called `prca` that performs predictable component analysis on the PCs. The homework assignment below will direct you to write this code in "pieces." In the end, you should combine all these pieces together to produce a single function that performs all the calculations. The header for this function should be:

```
prca.eof = function(data.eof,nens,ncon,alpha=0.01) {
## PERFORMAS PREDICTABLE COMPONENT ANALYSIS (OR EQUIVALENTLY MANOVA)
## ON PC-TIME SERIES OF AN ENSEMBLE FORECAST DATA SET
## INPUT:
##    DATA.EOF: LIST OUTPUT FROM EOF.LATLON
##    NENS: NUMBER OF ENSEMBLE MEMBERS
##    NCON: NUMBER OF CONDITIONS
##    ALPHA: SIGNIFICANCE LEVEL (DEFAULT = 1%)
## OUTPUT: LIST$
#    MIC: VALUES OF MUTUAL INFORMATION CRITERION
#    MIC.CRIT: SIGNIFICANCE VALUES OF MIC
#    F.MAX[NTRUN]: MAXIMIZED F-RATIOS
#    R[NENS,NCON,NTRUN]: PREDICTABLE VARIATES
#    P[NSPACE,NTRUN]: PREDICTABLE LOADING VECTORS
#    FEXPVAR[NMIN]: FRACTION OF VARIANCE EXPLAINED
#    RSQR.ADJ[NMIN]: ADJUSTED R-SQUARE
#    SNR.ADJ[NMIN]: ADJUSTED SIGNAL-TO-NOISE RATIO
#    NENS, NCON, NMIN, ALPHA: OTHER PARAMETERS
```

Exercise 18.1. Write an R program called `prca.null.fixdim` that estimates the 1% significance level of each eigenvalue from predictable component analysis using `ntrun` feature vectors. Run this program using `ntrun = 6` and state the significance values. The header of this program should be

```
prca.null.fixdim = function(ndim,ncon,nens,alpha=0.01,ntrial=10000) {
## ESTIMATES SIGNIFICANCE LEVEL OF THE ALL EIGENVALUES FROM PRCA/MANOVA
## FOR FIXED TRUNCATION NDIM
## BY MONTE CARLO METHODS
## INPUT:
#    NDIM: DIMENSION OF THE RANDOM VECTOR
#    NCOM: NUMBER OF CONDITIONS (E.G., NUMBER OF YEARS)
#    NENS: NUMBER OF ENSEMBLE MEMBERS (E.G., NUMBER OF REPITITIONS)
#    ALPHA: DESIRED SIGNIFICANCE LEVELS (CAN BE MORE THAN ONE)
#    NTRIALS: NUMBER OF MONTE CARLO TRIALS
## OUTPUT:
#    EVAL.CRIT[2,NDIM]: (alpha,1-alpha) SIGNIFICANCE THRESHOLDS FOR EACH EIGENVALUE
```

It is a good idea to start out with `ntrial = 100` while you are writing and debugging your code, and then change to `ntrial = 10000` at the last step.

□

Exercise 18.2. Write an R function to compute the MIC as a function of the number of PCs. The expression for MIC that is appropriate for PrCA is:

$$\text{MIC} = \log \left| \hat{\boldsymbol{\Sigma}}_N \right| - \log \left| \hat{\boldsymbol{\Sigma}}_T \right| + \frac{2CT + T(T+1)}{CE - C - T - 1} - \frac{2T + T(T+1)}{CE - 1 - T - 1}, \tag{18.1}$$

where T is the number of PCs. State the values of MIC that you obtain, and state the value of T at which the minimum occurs. □

Exercise 18.3. Write an R function to compute the maximized F-ratios from PrCA, and the associated adjusted signal-to-noise ratios and R-square. This function will need to compute the signal and noise covariance matrices from the PCs and then solve a generalized eigenvalue problem. State the maximized F-ratios for $T = 6$. Also state the adjusted signal-to-noise ratios and R-square. □

Exercise 18.4. Write an R function that computes the predictable variates. Remember to normalize the weight vector **q** so that the total variance equals unity. Run this function for $T = 6$. Make a plot of the variates for the two most predictable components. If the variate is dimensioned as `r[nens,nyrs,ntrun]`, then the variates for the most predictable component can be plotted using the command

```
boxplot(r[,,1],names=year,col='grey')
```

Label the plot as in the lecture notes, so that all information required to interpret the plot is contained in the plot. □

Exercise 18.5. Verify that the predictable variates are orthogonal. More generally, verify that the sample covariance matrix of the predictable components equals the identity matrix. □

Exercise 18.6. Write a function that calculates the loading vectors. Plot them. □