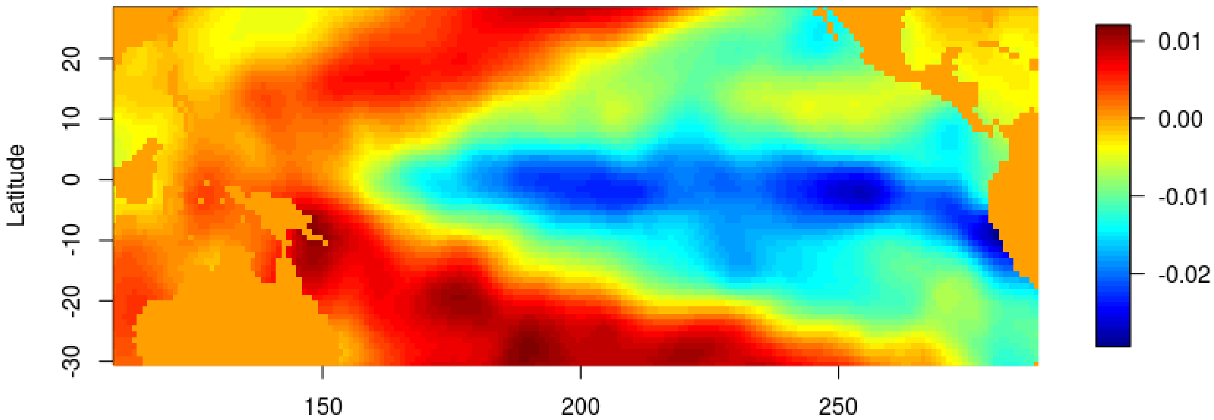
Assignment 4
Submitted By : Vikram Singh Chandel
Roll No. : 173040020

Q1

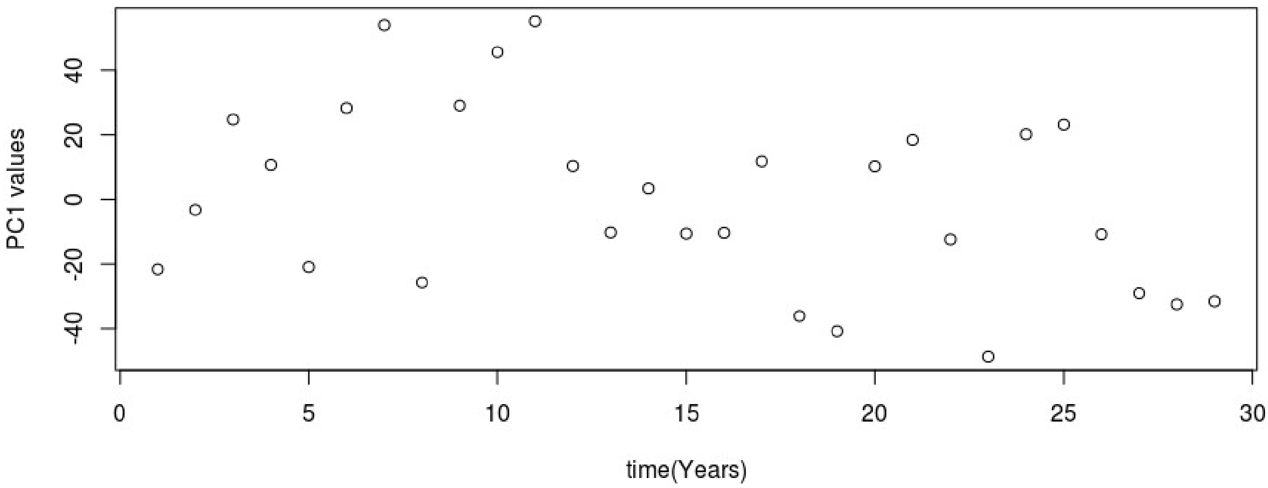
```
rm(list=ls())
setwd("/home/de/Dropbox/MTech/Hydroinformatics/Assignment 4/")
A=read.table('dataSST.txt',colClasses  = c('numeric','numeric','numeric','numeric'))
A[ ,1]=A[ ,1] - 108.5
A[ ,2]=A[ ,2] + 31.5
Temp=A[ ,4]
dim(Temp)=c(10800,29)

TEMPERATURE=as.data.frame(t(Temp))
PCA_Result=prcomp(TEMPERATURE,tol=0.0)
Lam=(PCA_Result$sdev)^2/sum((PCA_Result$sdev)^2)*100
CumLam=(cumsum(Lam))
# 1.i %variance explanied by first PC
pve=CumLam[1]

ev=PCA_Result$rotation
ev1=ev[ ,1]
dim(ev1)=c(180,60)
library(fields)
dev.new()
image.plot(seq(109.5,288.5),seq(-30.5,28.5),ev1,xlab='Longitude',ylab='Latitude',xlim=c(109.5,
288.5), ylim=c(-30.5, 28.5))
dev.off()
z1=PCA_Result$x[ ,1]
plot(z1,xlab='time(Years)',ylab='PC1 values')
dev.new()
barplot(CumLam,names.arg=seq(1:29),col='deepskyblue',xlab='No of Principlal
Pomponents',ylab='% variance explained')
lines(c(10,50),c(100,100),col='red',lw=2)
lines(c(10,50),c(90,90),col='green',lw=2)
legend('topleft',c('100%','90%'),lty=c(1,1),col=c('red','green'))
dev.off()

z1=PCA_Result$x[ ,1]
z2=PCA_Result$x[ ,2]

plot(z1,z2,xlab='First Principal Component',ylab='Second Principal Component')



which(CumLam>=90)
```

(i) The first principal component explains 50.13 % variance

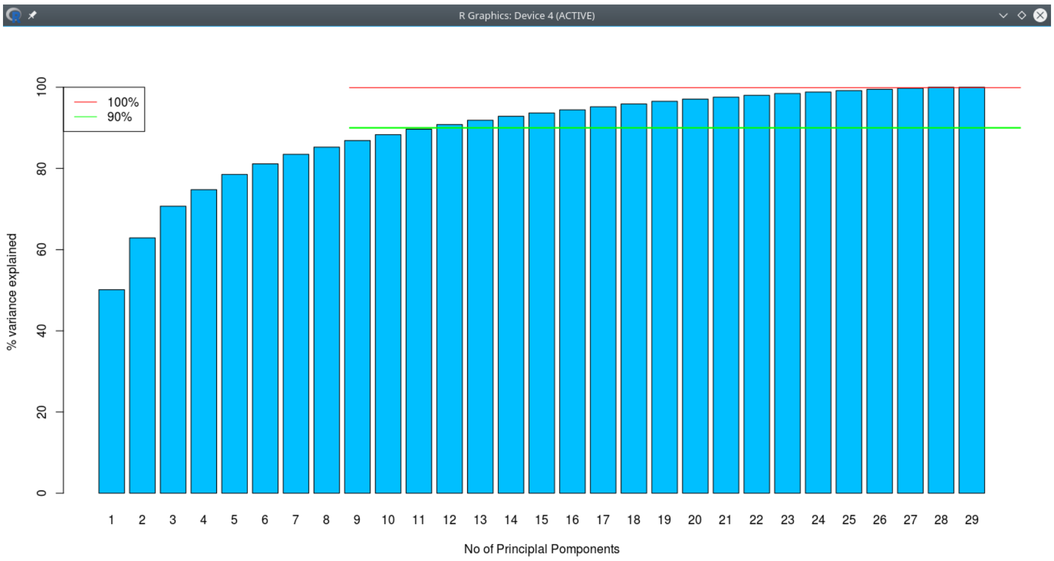
(ii) The following figure was obtained when first EOF was plotted spatially.The pattern is similar to ENSO.



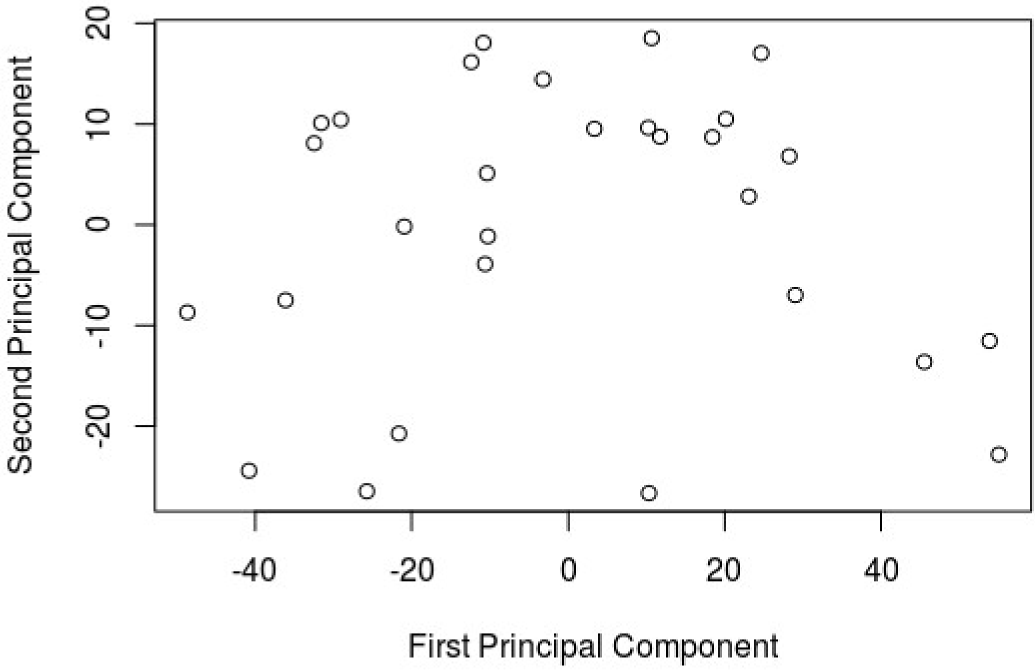
iii) The plot of first principal component with time.



iv)The bar chart of cumulative percent variance explained versus the number of Pcs show that the number of Pcs required toexplain 90% variance is 12.



v) The plot of first two principal components is shown in following figure.The data is uncorrelated which was the purpose of performing PCA.



Q2
i)

```
require(e1071)
FPI=0
dataCM=PCA_Result$x[ ,1:5]
coun=1
for(cc in 2:10)
{
 for(mm in seq(1.1,2,0.1))
{
 handle=cmeans (dataCM, cc, iter.max=100, dist="euclidean",method="cmeans", m=mm)
 F=sum((handle$membership)^2)/29;
 FPI[coun]=1-(cc*F-1)/(cc-1)
 coun=coun+1
}
}
# Choosing FPI close to 0.25
Ff=abs(FPI-0.25)
no=which.min(Ff)
cc=as.integer((no-1)/10)+2
mm=(no-10*(cc-2))*0.1+1

# from values of c= 3 and m= 1.4 the FPI is closest to 0.25
```

ii)

```
handle=cmeans (dataCM, cc, iter.max=100, dist="euclidean",method="cmeans", m=mm)
F=sum((handle$membership)^2)/29;
FPIf=1-(cc*F-1)/(cc-1)
# 0.2534342
MembMat=handle$membership
```

Q3

i)

```
RAW=read.xlsx('dataCCA.xlsx',1,colClasses = c('numeric','numeric','numeric','numeric','numeric'))
RAW[1,1]=0.00
names(RAW)=c('prec','conv','olr','slp','theta')
Y=RAW[1:480,c(1,2)]
X=RAW[1:480,c(4,5,3)]
CCH=cc(X,Y)
summary(CCH)
# first and second correlations
CCH$cor
plot(CCH$scores$xscores[,1],CCH$scores$yscores[,1],xlab='U1',ylab='V1')
```

ii) The magnitudes of the two correlations are 0.584 and 0.286

iii)Plot of the first canonical variate pair is shown below. There is some correlation between them a it can be seen from the scatter plot. The coefficient of theta ie Surface Potential Temerature had coeff of 0.22 (this was only positive value among other coeficients) in first canonical component. So it might have a positive influence on the linear combination of ppt and conv_ppt. The reason for this may be that with increasing temperature evapotranpiration increases and which may subsequently lead to precipitaion.