Classifying radar waveforms from the satellite Topex/Poseidon

- **Statistical aim**

  We recall that for each unit i, we observe one waveform (i.e. curve) measured at 70 points (i.e. $\boldsymbol{x}_i = (\chi_i(t_1), \ldots, \chi_i(t_{70}))$. The file "npfda-sat.dat" contains the observed curves $\{\boldsymbol{x}_i\}_{i=1,\ldots,472}$. This case study is not presented in the NPFDA book and for more details, see Dabo-Niang, Ferraty and Vieu (2006, Far East J. Theor. Stat., 18, 93-119).

- **R/S+ commandlines**:

  - **Entering spectrometric data**

    ```
    SATDAT <- as.matrix(read.table("npfda-sat.dat"))
    attributes(SATDAT)$dimnames <- NULL
    ```

  - **Automatical classification: the following commandline allows to get directly the groups**:

    ```
    sat.classif <- classif.npfda(SATDAT, 1:ncol(SATDAT),
                    kind.of.kernel="quadratic", semimetric="hshift",
                    threshold=0.05, nb.bw=50, nss=0, mspg=10,
                    centrality="median")
    ```

    Remark: `nss=0` means that the criterion HI is used instead of SHI whereas `centrality="mean"` implies that HI is based on differences between median and modal curves.

  - **Loading the groups**

    ```
    Group11 <- sat.classif$Partition[[1]]
    Group12 <- sat.classif$Partition[[2]]
    Group21 <- sat.classif$Partition[[3]]
    Group22 <- sat.classif$Partition[[4]]
    Group23 <- sat.classif$Partition[[5]]
    ```

    Remark: `spec.classif$Labels` contains the history of the splitting procedure namely '11', '12' '21', '22' and '23'.

- **Plotting terminal leaves: Group11, Group12, Group21, Group22 and Group23**
  - **Displaying a sample of 16 waveforms for each groups**
    The following R/S+ commandlines allow to build Figure 1:

```
s11 <- sample(Group11,16)
s12 <- sample(Group12,16)
s21 <- sample(Group21,16)
s22 <- sample(Group22,16)
s23 <- sample(Group23,16)
x <- 1:ncol(SATDAT)
close.screen(all = TRUE)
screen1 <- c(0.14,0.41,0.55,0.88)
screen2 <- c(0.58,0.85,0.55,0.88)
screen3 <- c(0,0.32,0.11,0.44)
screen4 <- c(0.34,0.66,0.11,0.44)
screen5 <- c(0.68,1,0.11,0.44)
split.screen(rbind(screen1, screen2, screen3, screen4, screen5))
screen(1)
par(mar=c(0,0,0,0)+0.2, oma=c(0,0,2,0), xaxt="n", yaxt="n")
plot(x,x,type="n",bty="n")
mtext(c("GROUP 11","GROUP 12"), side=3, outer=T, line=-4,
      at=c(0.28,.72))
##
# GROUP11
#########
nbscreen <- split.screen(c(4,4),1)
screen(nbscreen[1])
plot(x, SATDAT[s11[1],],ylim=range(SATDAT), type="l",
     xlab="",ylab="")
count <- 0
for(i in s11[2:16]){
        count <- count + 1
        screen(nbscreen[count+1])
        plot(x,SATDAT[i,],ylim=range(SATDAT), type="l",
             xlab="",ylab="")
}
mtext(c("GROUP 21","GROUP 22", "GROUP 23"), side=3, outer=T,
```

```
      line=-22, at=c(0.16,0.5,0.84))
##
# GROUP12
#########
nbscreen <- split.screen(c(4,4),2)
screen(nbscreen[1])
plot(x, SATDAT[s12[1],],ylim=range(SATDAT), type="l",
     xlab="",ylab="")
count <- 0
for(i in s12[2:16]){
        count <- count + 1
        screen(nbscreen[count+1])
        plot(x,SATDAT[i,],ylim=range(SATDAT), type="l",
             xlab="",ylab="")
}
##
# GROUP21
#########
nbscreen <- split.screen(c(4,4),3)
screen(nbscreen[1])
plot(x, SATDAT[s21[1],],ylim=range(SATDAT), type="l",
     xlab="",ylab="")
count <- 0
for(i in s21[2:16]){
        count <- count + 1
        screen(nbscreen[count+1])
        plot(x,SATDAT[i,],ylim=range(SATDAT), type="l",
             xlab="",ylab="")
}
##
# GROUP22
#########
nbscreen <- split.screen(c(4,4),4)
screen(nbscreen[1])
plot(x, SATDAT[s22[1],],ylim=range(SATDAT), type="l",
     xlab="",ylab="")
count <- 0
for(i in s22[2:16]){
```

```
        count <- count + 1
        screen(nbscreen[count+1])
        plot(x,SATDAT[i,],ylim=range(SATDAT), type="l",
             xlab="",ylab="")
}
##
# GROUP23
#########
nbscreen <- split.screen(c(4,4),5)
screen(nbscreen[1])
plot(x, SATDAT[s23[1],],ylim=range(SATDAT), type="l",
     xlab="",ylab="")
count <- 0
for(i in s23[2:16]){
        count <- count + 1
        screen(nbscreen[count+1])
        plot(x,SATDAT[i,],ylim=range(SATDAT), type="l",
             xlab="",ylab="")
}
```

•• **Loading mean and modal curves for each terminal group**

```
MODAL.CURVES <- spec.classif$MODES
MEAN.CURVES <- spec.classif$MEAN
```

•• **Displaying modal curves for each terminal group**

The following commandlines allow to perform Figure 2:

```
par(mfrow=c(1,6), mar=c(1,2,2,0)+.2, oma=c(1,1,0,0), pty="s")
Labels <- sat.classif$Labels
for(j in 1:2)
    plot(x, MODAL.CURVES[j,],ylim=range(SATDAT), type="l",
         xlab="",ylab="", main=paste("GROUP",Labels[j],sep=""),
         cex.main=2)
plot(x, MODAL.CURVES[j,],ylim=range(SATDAT), type="n", xlab="",
     ylab="",bty="n", xaxs=F, yaxs=F)
for(j in 3:5)
    plot(x, MODAL.CURVES[j,],ylim=range(SATDAT), type="l", xlab="",
         ylab="", main=paste("GROUP",Labels[j],sep=""), cex.main=2 )
```

•• **Displaying mean curves for each terminal group**

The following commandlines allow to perform Figure 3:

```
par(mfrow=c(1,6), mar=c(1,2,2,0)+.2, oma=c(1,1,0,0), pty="s")
Labels <- sat.classif$Labels
for(j in 1:2)
    plot(x, MEAN.CURVES[j,],ylim=range(SATDAT), type="l", xlab="",
         ylab="", main=paste("GROUP",Labels[j],sep=""), cex.main=2)
plot(x, MEAN.CURVES[j,],ylim=range(SATDAT), type="n", xlab="",
     ylab="",bty="n", xaxs=F, yaxs=F)
for(j in 3:5)
    plot(x, MEAN.CURVES[j,],ylim=range(SATDAT), type="l", xlab="",
         ylab="", main=paste("GROUP",Labels[j],sep=""), cex.main=2 )
```

•• **Computing mean and median curves for Groups 1 and 2**

```
nb.groups <- length(sat.classif$Partition)
MEAN.CURVES <- matrix(0, 2, ncol(SATDAT))
MEDIAN.CURVES <- matrix(0, 2, ncol(SATDAT))
MODAL.CURVES <- matrix(0, 2, ncol(SATDAT))
Group1 <- c(sat.classif$Partition[[1]],
            sat.classif$Partition[[2]])
Group2 <- c(sat.classif$Partition[[3]],
            sat.classif$Partition[[4]],
            sat.classif$Partition[[5]])
)
MEAN.CURVES[1,] <- apply(SATDAT[Group1,], 2, mean)
MEAN.CURVES[2,] <- apply(SATDAT[Group2,], 2, mean)
MEDIAN.CURVES[1,] <- SATDAT[Group1[median.npfda(SEMIMETRIC
                            [Group1, Group1])],]
MEDIAN.CURVES[2,] <- SATDAT[Group2[median.npfda(SEMIMETRIC
                            [Group2, Group2])],]
```

•• **Computing modal curves for groups 1 and 2**

```
grid <- 1:ncol(SATDAT)
SM.SAT <- semimetric.hshift(SATDAT, SATDAT, grid)
Hrange <- range(SM.SAT)
Bw.seq <- seq(Hrange[1], Hrange[2] * 0.5, length = 50 + 1)[-1]
PROBCURVES <- matrix(0, n, 50)
for(i in 1:n){
        PROBCURVES[i,] <- prob.curve(i, SM.SAT, Bw.seq)
}
res.classif.bw <- classif.bw(PROBCURVES, Bw.seq, 1:n)
```

```
Bw.opt <- res.classif.bw$bw
index <- res.classif.bw$index
shi <- classif.hi(SATDAT, SM.SAT, "quadratic", Bw.opt, 1:n,
                  "hshift", "median", grid)
first.split.sat <- classif.part(SATDAT, PROBCURVES, SM.SAT,
                                "quadratic", index, Bw.seq, 1:n,
                                shi, "hshift", 0.05, 0, 10,
                                "median", grid)
Group1 <- first.split.sat$Groups[[1]]
Group2 <- first.split.sat$Groups[[2]]
Band <- first.split.sat$Bw.opt
rank.mode1 <- funopa.mode(Band[1], SM.SAT[Group1,Group1],
                          "quadratic")
MODAL.CURVES[1,] <- CURVES[Group1[rank.mode1],]
rank.mode2 <- funopa.mode(Band[2], SM.SAT[Group2, Group2],
                          "quadratic")
MODAL.CURVES[2,] <- CURVES[Group2[rank.mode2],]
```

•• **Displaying modal, median and mean curves for groups 1 and 2**

The following commandlines allow to perform Figure 4:

```
par(mfrow=c(2,3), mar=c(1,2,2,0)+.2, oma=c(1,1,0,0), pty="s")
for(j in 1:2){
    plot(x, MODAL.CURVES[j,],ylim=range(SATDAT), type="l",
         xlab="",ylab="", main=paste("Modal curve of GROUP",j,sep=""),
         cex.main=1.5)
    plot(x, MEDIAN.CURVES[j,],ylim=range(SATDAT), type="l", xlab="",
         ylab="", main=paste("Median curve of GROUP",j,sep=""),
         cex.main=1.5)
    plot(x, MEAN.CURVES[j,],ylim=range(SATDAT), type="l", xlab="",
         ylab="", main=paste("Mean curve of GROUP",j,sep=""),
         cex.main=1.5)
}
```

•• **Displaying the splitting score behavior along the procedure**

The following commandlines allow to perform Figure 5:

```
Splitting.score <- sat.classif$Ssc
Split.names <- c(paste("GROUPS","\n","1 & 2"),
                 paste("GROUPS", "\n","11 & 12"),
```

```
                  paste("GROUPS", "\n","21, 22 & 23"))
par(mfrow=c(1,1))
barplot(Splitting.score, names=Split.names,
        main="Splitting scores", ylab="Percentages")
abline(h=0.05, lwd=3)
```

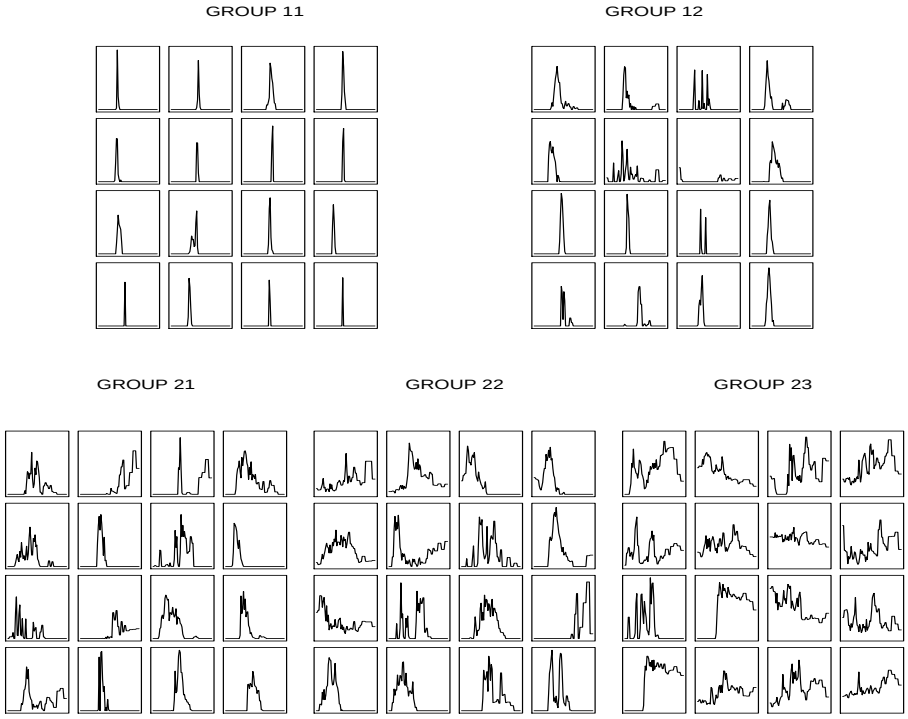

Figure 1: Samples of radar waveforms for each group

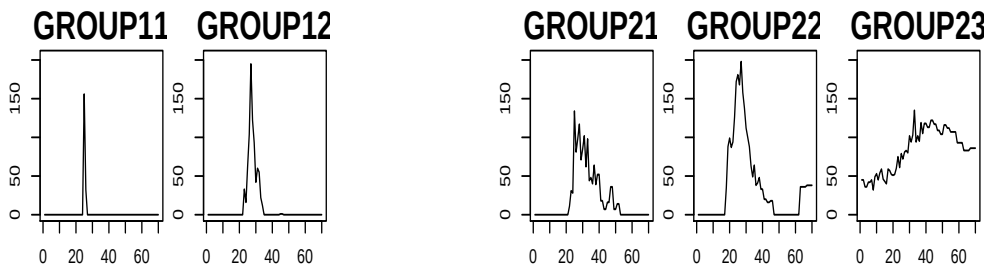

Figure 2: Modal curves for the five terminal groups

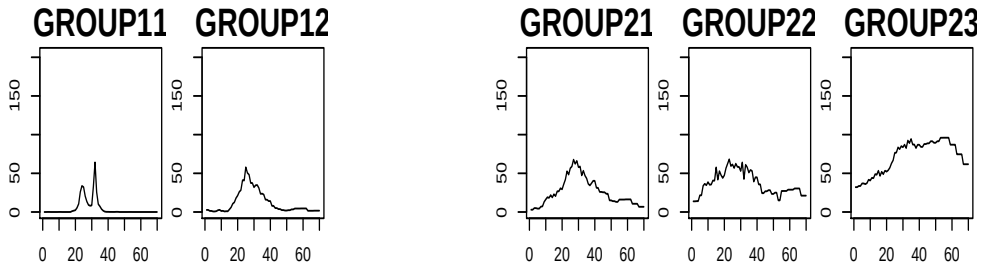

Figure 3: Mean curves for the five terminal groups

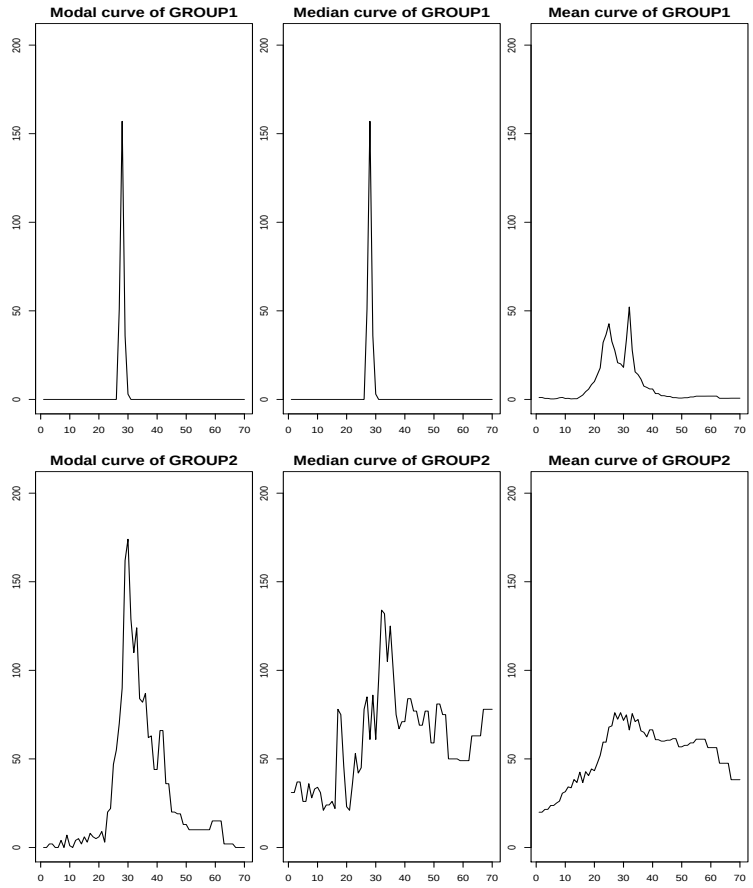

Figure 4: First splitting into $GROUP\ 1$ and $GROUP\ 2$: comparison between modal, median and mean curves

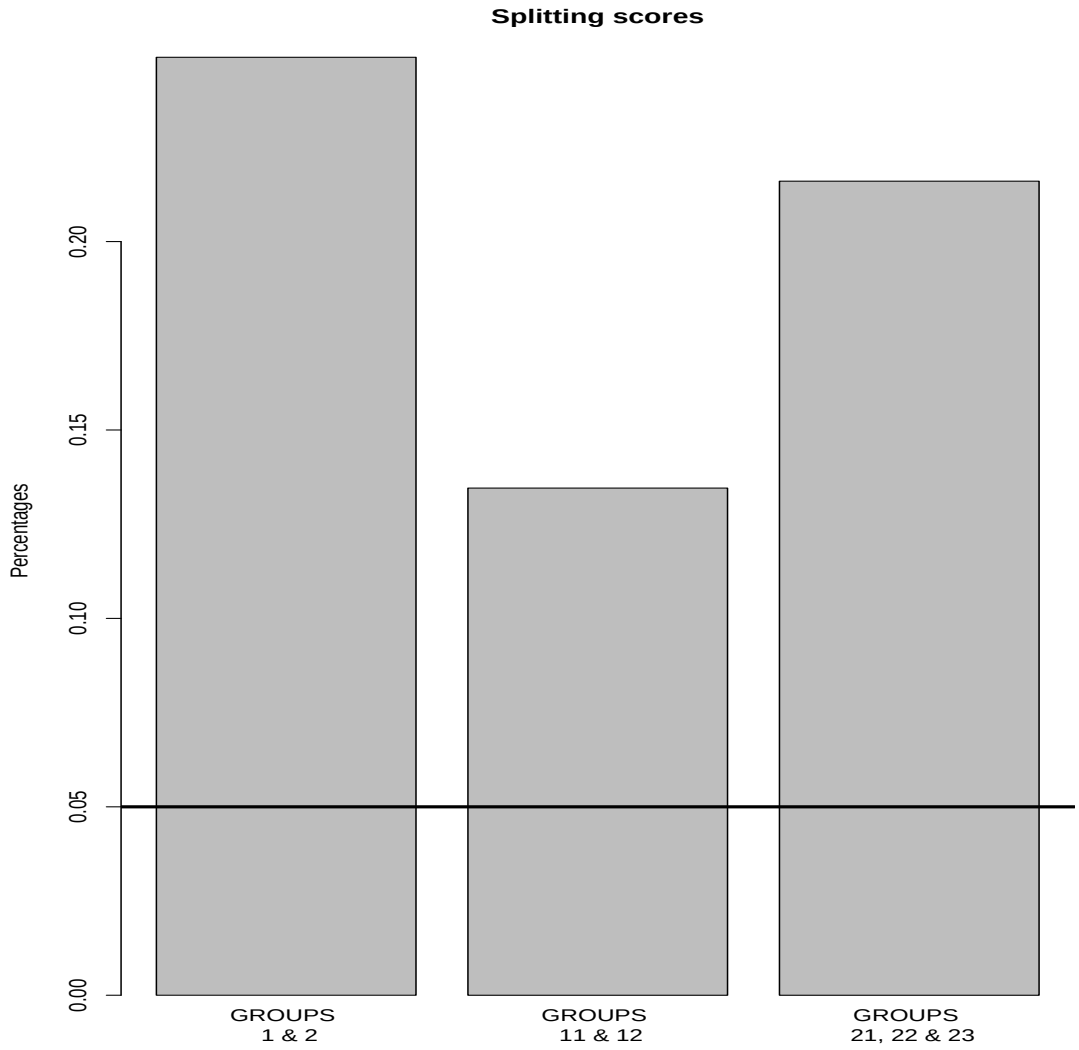

Figure 5: Behavior of splitting score for the satellite waveforms; the horizontal line corresponds to the threshold of splitting score