Nonparametric Analysis of US Dairy Production and Consumption

Functional Depth on clusters

Teo Bucci* Filippo Cipriani† Gabriele Corbo‡ Andrea Puricelli§

2023-02-17

Contents

1 Load libraries and data

```
library(BNPTSclust)
library(roahd)
library(fda.usc)
```

```
data_path = file.path('data_updated_2021')
output_path = file.path('output')
dairy = read.table(file.path(data_path, 'dairy.csv'),header = T,sep = ';')
```

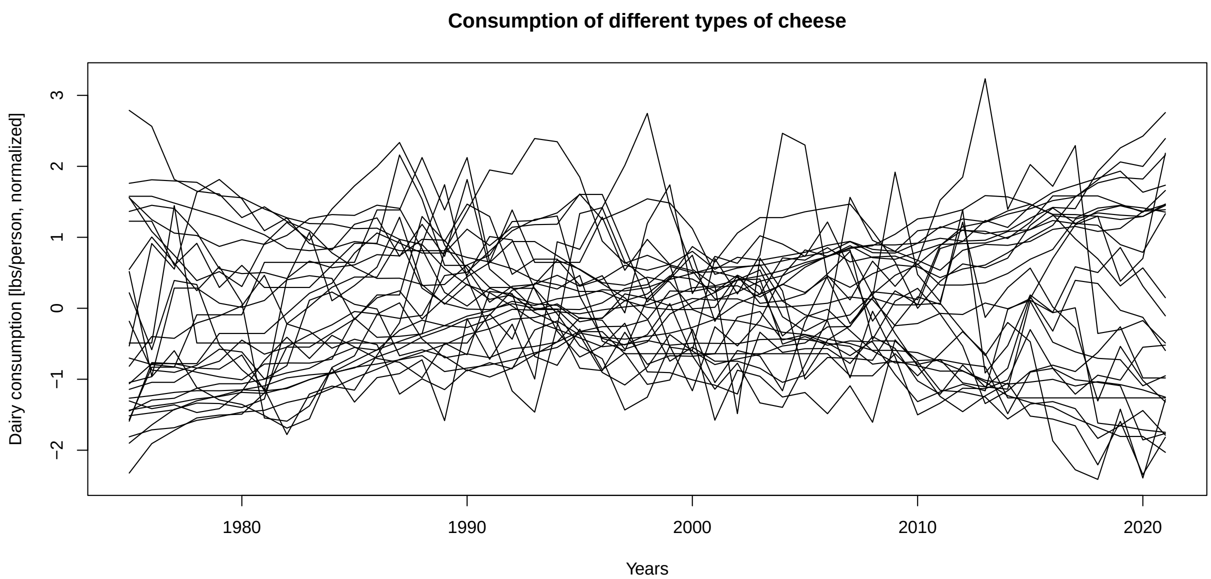
Remove total cheese consumptions:

```
dairy <- dairy[,-c(12,13,14,15)]
```

2 Bayesian nonparametric clustering

*teo.bucci@mail.polimi.it
†filippo.cipriani@mail.polimi.it
‡gabriele.corbo@mail.polimi.it
§andrea3.puricelli@mail.polimi.it

```
years = 1975:2021
matplot(
    years,
    dairy,
    type = 'l',
    lty = 1,
    col = "black",
    main = "Consumption of different types of cheese",
    xlab = "Years",
    ylab = "Dairy consumption [lbs/person, normalized]"
)
```

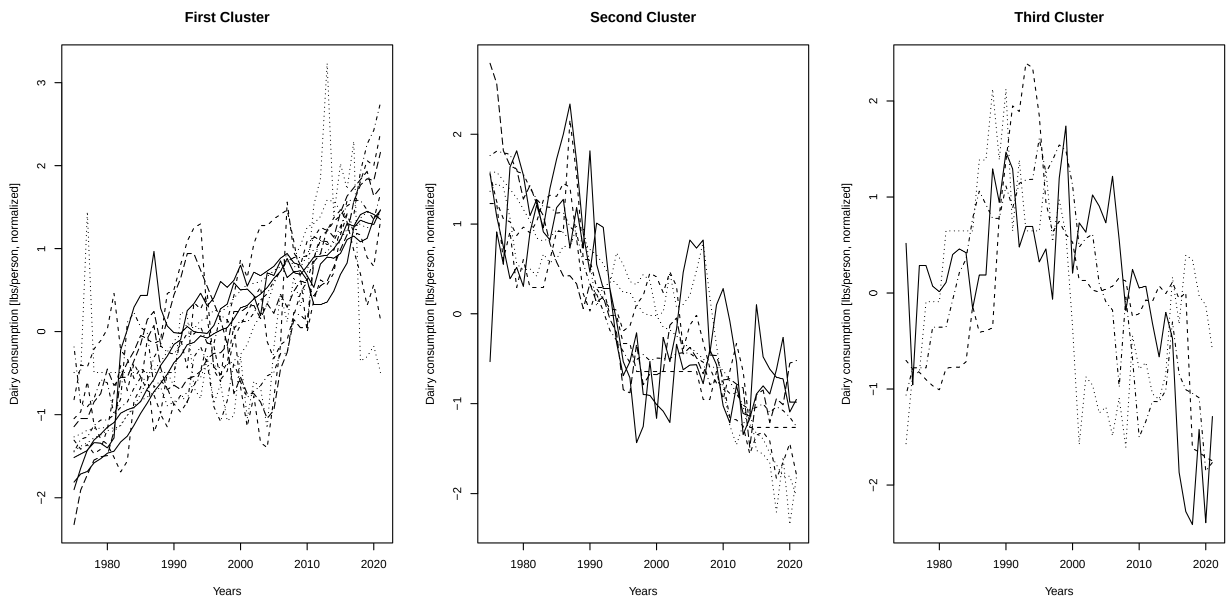


Perform the clustering using the function `tseriescm` from the package `BNPTSclust`, based on a nonparamteric bayesian approach.

```
tseriescm.out <-
    tseriescm(
        dairy,
        maxiter = 100,
        burnin = 10,
        thinning = 2,
        level = FALSE,
        trend = TRUE,
        seasonality = FALSE,
        priorb = TRUE,
        b = 0
    )
```

The clustering obtains five clusters, three of which non singular:

```r
par(mfrow = c(1, 3))
matplot(
    years,
    dairy[, tseriescm.out$gnstar == 1],
    type = 'l',
    col = "black",
    xlab = 'Years',
    ylab = 'Dairy consumption [lbs/person, normalized]',
    main = "First Cluster"
)
matplot(
    years,
    dairy[, tseriescm.out$gnstar == 2],
    type = 'l',
    col = "black",
    xlab = 'Years',
    ylab = 'Dairy consumption [lbs/person, normalized]',
    main = 'Second Cluster'
)
matplot(
    years,
    dairy[, tseriescm.out$gnstar == 4],
    type = 'l',
    col = "black",
    xlab = 'Years',
    ylab = 'Dairy consumption [lbs/person, normalized]',
    main = 'Third Cluster'
)
```



3 Functional Depth measures on clusters

```
c1 = dairy[, tseriescm.out$gnstar == 1]
c2 = dairy[, tseriescm.out$gnstar == 2]
c3 = dairy[, tseriescm.out$gnstar == 4]
grid = 1:dim(dairy)[1]
c1 = as.matrix(c1)
c1_f = fData(grid, t(c1))
c2 = as.matrix(c2)
c2_f = fData(grid, t(c2))
c3 = as.matrix(c3)
c3_f = fData(grid, t(c3))
```

Compute the modified band depth

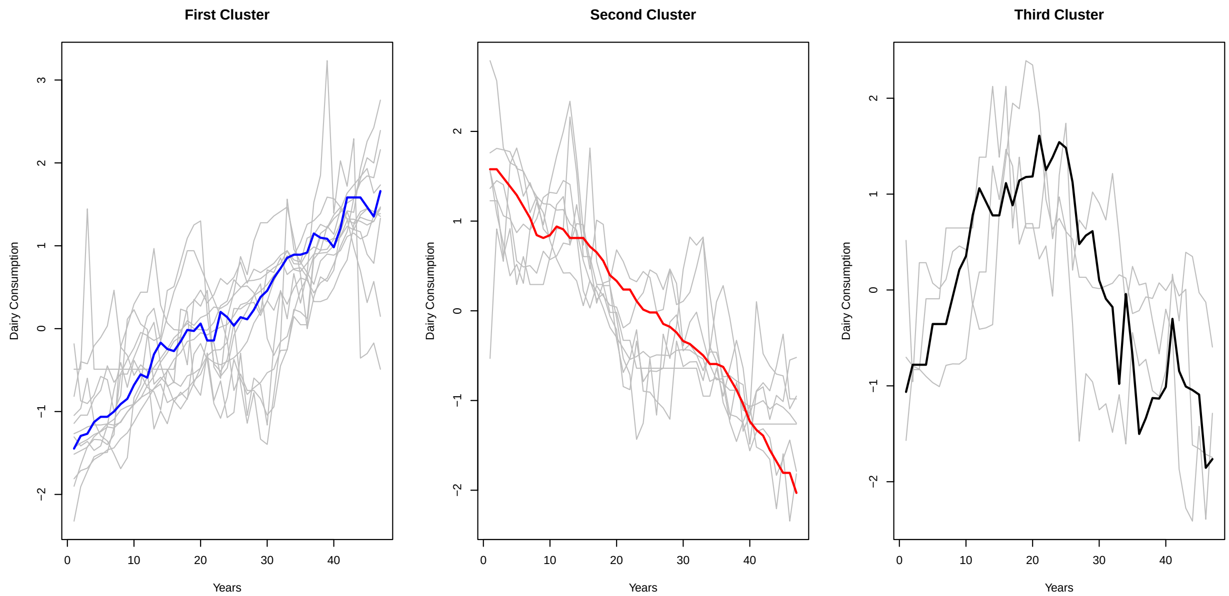
```
median_curve1 <- median_fData(fData = c1_f, type = "MBD")
median_curve2 <- median_fData(fData = c2_f, type = "MBD")
median_curve3 <- median_fData(fData = c3_f, type = "MBD")
```

```
par(mfrow = c(1, 3))

plot(c1_f, col = "grey", xlab = "Years", ylab = "Dairy Consumption", main = "First Cluster")
grid_ecg1 <- seq(median_curve1$t0, median_curve1$tP, by = median_curve1$h)
lines(grid_ecg1, median_curve1$values, col = "blue", lwd = 2)

plot(c2_f, col = "grey", xlab = "Years", ylab = "Dairy Consumption", main = "Second Cluster")
grid_ecg2 <- seq(median_curve2$t0, median_curve2$tP, by = median_curve2$h)
lines(grid_ecg2, median_curve2$values, col = "red", lwd = 2)

plot(c3_f, col = "grey", xlab = "Years", ylab = "Dairy Consumption", main = "Third Cluster")
grid_ecg3 <- seq(median_curve3$t0, median_curve3$tP, by = median_curve3$h)
lines(grid_ecg3, median_curve3$values, col = "black", lwd = 2)
```

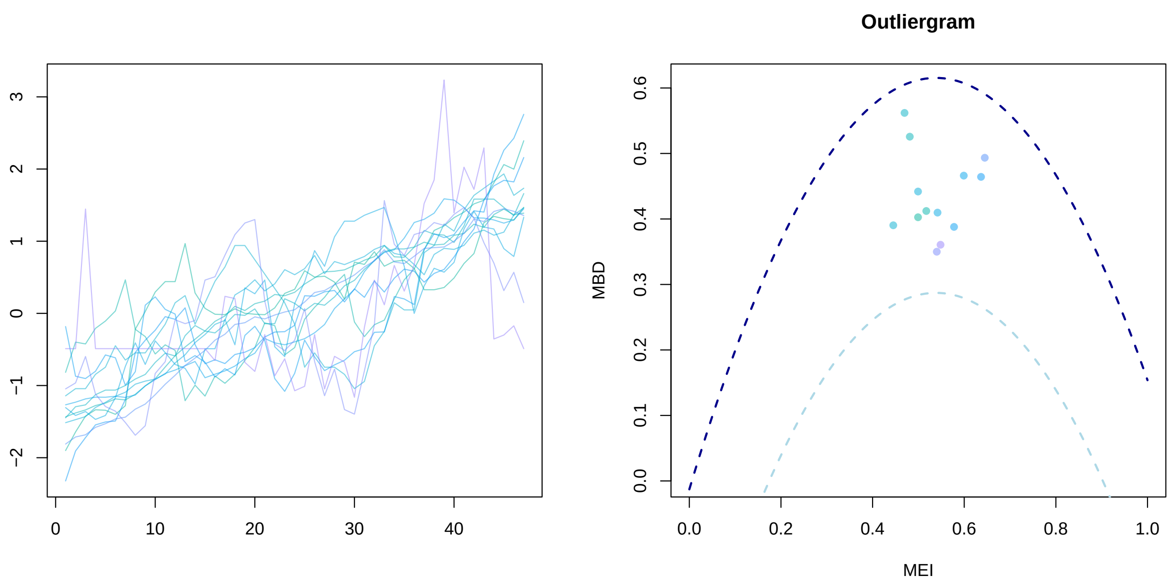


We can note an up tren ecc.....

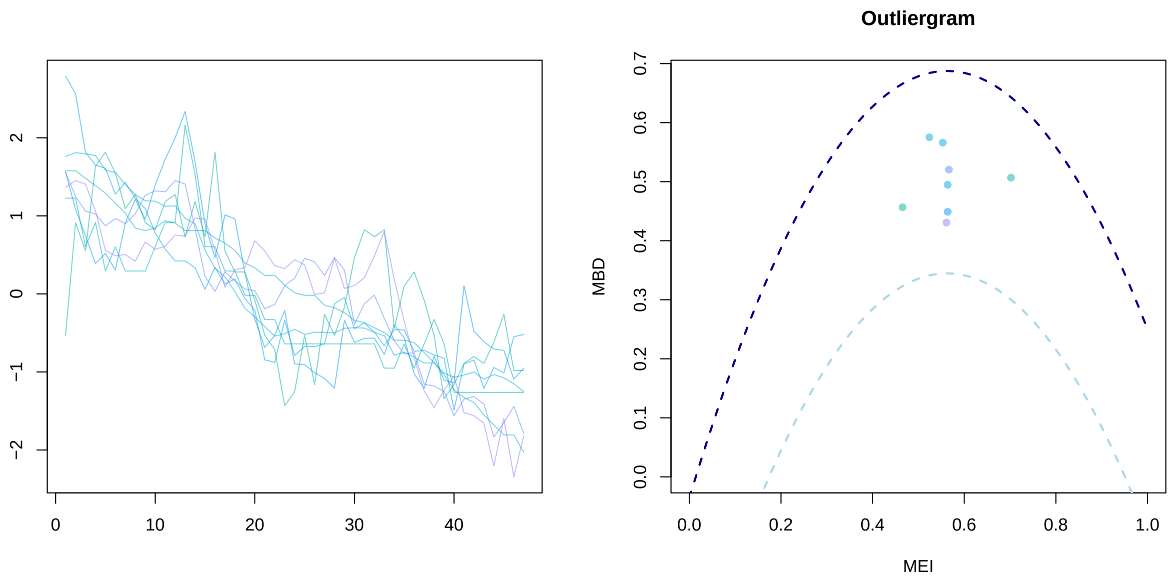
```
par(mfrow=c(1,3))
invisible(fbplot(c1_f))
invisible(fbplot(c2_f))
invisible(fbplot(c3_f))
```

There are no outliers in the 3 clusters.

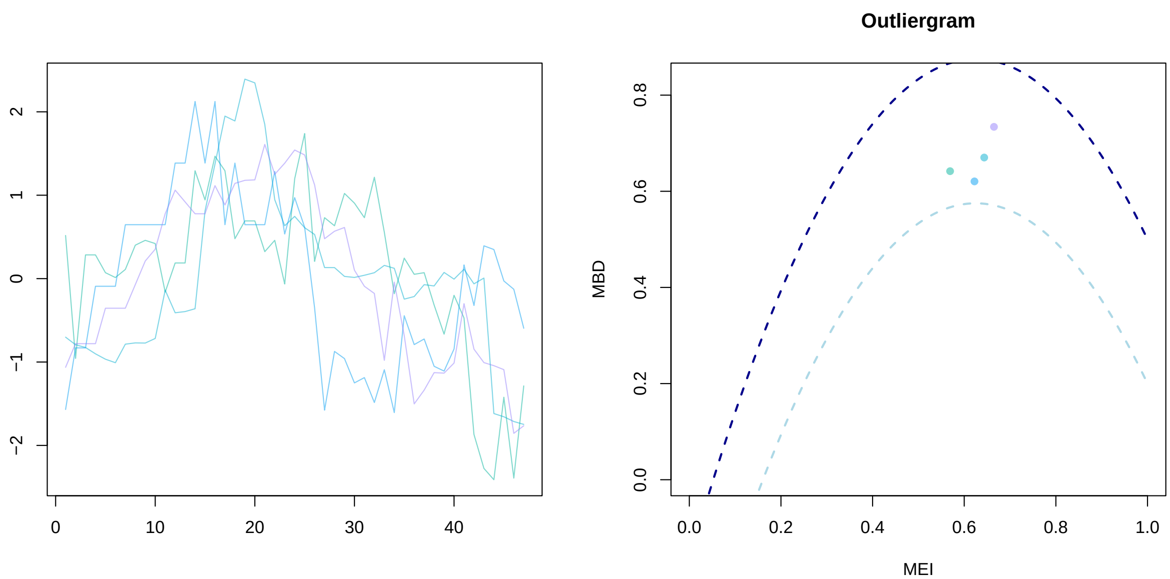
```
invisible(outliergram(c1_f))
```



```
invisible(outliergram(c2_f))
```




```
invisible(outliergram(c3_f))
```



4 Cluster analysis

```r
c1_names <- colnames(c1)
knitr::kable(c1_names, col.names="Cluster 1")
```

Cluster 1
Cheddar
American_Other
Mozzarella
Italian_other
Muenster
Cream_and_Neufchatel
Other_Dairy_Cheese
fluid_yogurt
butter
cheese_american
cheese_other
evap_cnd_bulk_and_can_skim_milk
dry_buttermilk

```r
c2_names <- colnames(c2)
knitr::kable(c2_names, col.names="Cluster 2")
```

Cluster 2
Swiss
Brick
fluid_milk
cheese_cottage
evap_cnd_canned_whole_milk
evap_cnd_bulk_whole_milk
frozen_ice_cream_regular
frozen_sherbet

```r
c3_names <- colnames(c3)
knitr::kable(c3_names, col.names="Cluster 3")
```

Cluster 3
Foods_and_spreads
frozen_other
dry_whole_milk
dry_whey