flowFDA: Functional Data Analysis for Flowcytometry Data

Lieven Clement & Olivier Thas

February 21, 2014

Contents

1 Introduction

The `flowFDA` package can be used for analysing flow cytometry (FC) experiments with functional model based clustering, functional principal component analysis, functional discriminant analysis and to compare multivariate flowcytometry fingerprints across treatments.

Flow cytometry (FC) can generate fast fingerprints by characterizing the multivariate distribution of cellular features of single cells. We developed a statistical pipeline for classifying samples and for inferring on distributional changes induced by experimental factors. Our method consists of 1) Creating a quantitative fingerprint from the multivariate distribution, 2) Extracting informative fingerprint features by discriminant analysis, 3) Permutation tests for assessing differences across treatment groups in the reduced feature space and 4) Interpreting these differences in terms of changes in the multivariate FC distribution. We illustrate our method on a case study, which aims at detecting changes in microbial community composition of drinking water induced by environmental stress.

The example data used in this vignette are a subset of the data provided by De Roy et al. (2012). It contains a flowset of n=30 different flows for the stress experiment. Two types of treatments were conducted on Evian water to simulate changing physico- chemical conditions: temperature and nutrient treatment. For the heat treatment, 1 L bottles were incubated for 3 and 24 h at 37 degrees Celsius. For the nutrient treatment, 1 mL of water was replaced by 1 mL of a 1/10 dilution of autoclaved Luria-Bertani broth (10 g tryptone, 5 g yeast extract and 10 g NaCl per L) to a final TOC of 0.65 mg/L. The bottles were incubated for 3 and 24 h at room temperature. The five treatments are coded as follows: control (c), 3h heat treatment (h3), 24h heat treatment (h24), 3h nutrient treatment (n3) and 24h nutrient treatment (n24). Two fluorescent dyes, SYBR Green and Propidium Iodide, were used in combination as a viability indicator. The channels SS Log, FL 1 Log and FL 3 Log are used in the vignette, which correspond to the bandwidth filters for the side scatter and the staining.

2 Importing Data

The package builds upon the `flowCore` package to import raw flow cytometric data files in R (Ellis et al., 2013). We first have to load the `flowFDA` package and read the flowset. The `read.flowSet()` function will read all fcs files in the directory specified in `path` argument. In our pipeline, informative file names are used, i.e. treatmentx_replicatey.fcs. This accommodates a straightforward automation of the data analysis workflow. Users that want to import their own FC experiments can modify the commented code below.

```
> library(flowFDA)
> #fset<-read.flowSet(path="~/Dropbox/LabMet/flowcytometry/stress_test_2/",
> #transformation=FALSE)
> #fset
> #
> ##subset feet to reduce memory footprint
> #param=c("SS Log","FL 1 Log","FL 3 Log")
> #fset=fset[,param]
> #fset
```

We will use the channels SS Log, FL 1 Log and FL 3 Log, which correspond to the side scatter, SYBR green and Propidium Iodide staining bandpass filters. The data have been transformed to fall within a range of 0 and 1 and extremely low intensities are removed using a rectangular gate so as to avoid artefacts. The flowcytometer used in this study returned log transformed intensities and had a maximum log transformed intensity of 2^{16}.

```
> #mytrans<-function(x) x/2^16
> #fset<-transform("FL 1 Log"=mytrans,"FL 3 Log"=mytrans,"SS Log"=mytrans)%on%fset
> #rg <- rectangleGate(filterId="myRectGate", list("SS Log"=c(1/2^17, Inf),
> #"FL 1 Log"=c(1/2^17, Inf),"FL 3 Log"=c(1/2^17,Inf)))
> #fset<-Subset(fset,rg)
```

For other flow cytometers, it might be nessecary to log-transform the data first. This can be done by adopting a customized transform function or by setting the transformation flag in the `read.flowSet` function.

```
> #logtrans<- function(x) log(x)
```

A good choice of the filename can enable an automated construction of the grouping variable

```
> #construct experiment factor
> #files<-list.files(path="~/Dropbox/LabMet/flowcytometry/stress_test_2/",pattern=".fcs")
> #expHlp<-unlist(strsplit(files,split="_replicate"))
> #dim(expHlp)<-c(2,length(fset))
> #group<-as.factor(expHlp[1,])
> #nGroup<-nlevels(group)
```

The steps above have been performed on the example data. The data has not been integrated in the `flowFDA` package to comply with the Bioconductor guidelines with regard to the size of software packages. The example dataset can be found in the `flowFDAExampleData` Bioconductor data package and can be downloaded in the usual way.

```
> source("http://www.bioconductor.org/biocLite.R")
> biocLite("flowFDAExampleData")
```

First of we load the package and the experiment data package:

```
> library(flowFDAExampleData)
> library(flowFDA)
```

```
> data(fset)
> data(group)
> param=c("SS Log","FL 1 Log","FL 3 Log")
> nGroup=nlevels(group)
> nSamp=length(fset)
> groupLevels=levels(group)
```

3 Fingerprinting: constructing flowBasis object

A `flowBasis` object contains fingerprints of the multivariate flow cytometry (FC) distributions of the N different samples in the study. The package default constructs a fingerprint using all $q = (n_{\text{channel}}) \times (n_{\text{channel}} + 1)/2$ pairwise bivariate densities of the n_{channel} flow channels of interest. Each bivariate density is estimated using a kernel density estimator with bandwidth bw. Next, functional data analysis in the `flowFDA` package is provided by a discretisation approach: i.e. the kernel density estimators are used to calculate smoothed density estimates on an equally spaced mesh with $n_{\text{bin}} \times n_{\text{bin}}$ grid points. This yields an $N \times r$ data matrix $\boldsymbol{X}$ (with $r = q \times n_{\text{bin}}^2$) that can be processed with standard multivariate data analysis methods (e.g. Ramsay and Silverman, 2005). The rows of the matrix $\boldsymbol{X}$ are referred to as the fingerprints. The fingerprints thus provide a straightforward graphical interpretation in terms of the bivariate distributions of flow channels.

```
> fbasis=flowBasis(fset,param,nbin=128, bw=0.01)
> fbasis
```

```
flowBasis object
Kernel Density Estimation on grid of 128 x 128
```

```
Kernel density bandwith: 0.01

bivariate densities for channels
     [,1]       [,2]
[1,] "SS Log"   "FL 1 Log"
[2,] "SS Log"   "FL 3 Log"
[3,] "FL 1 Log" "FL 3 Log"
```

Similar to De Roy et al. (2012), one might also use probability binning (PB) for fingerprinting. Details are provided in the `flowFDAProbabilityBinning` vignette. However, interpretation based on plots using PB is less convenient.

4 Data exploration

The bivariate densities can be explored graphically. An example for the third flowset can be generated using the code below. The plot for the kernel density basis is given in Figure 1.

```
> par(mfrow=c(2,2))
> plot(fbasis,ask=FALSE,samples=3)
```

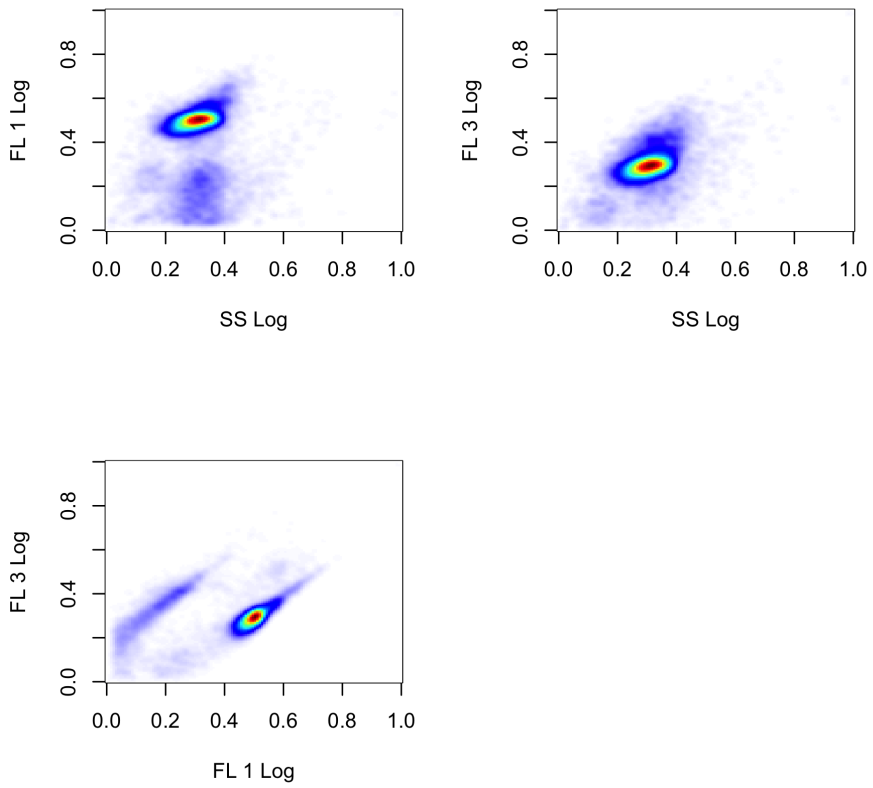

Figure 1: Plot of bivariate kernel density fingerprint for third flow of the flowSet. White regions correspond with regions of very low density, the blue-to-red gradient corresponds to low-to-high densities.

The fingerprints can also be averaged over several flows, e.g. for the flows belonging to the same group. An example of the graphical interpretation of the averaged fingerprint for the control group (c) and 24h nutrient treatment group (n24) are given in Figure 2. The cells for the c-group are more tightly centered around a SS of .35, FL1 of .5 and a FL3 of 0.3. For the n24-group the distribution has a larger variance and includes a considerable number of cells at higher SS, FL1 and FL3 values. This region corresponds to larger cells with more nucleic acids and intact membranes.

```
> par(mfrow=c(2,3))
> plot(fbasis,ask=FALSE,samples=group==groupLevels[1],main=groupLevels[1])
> plot(fbasis,ask=FALSE,samples=group==groupLevels[4],main=groupLevels[4])
```

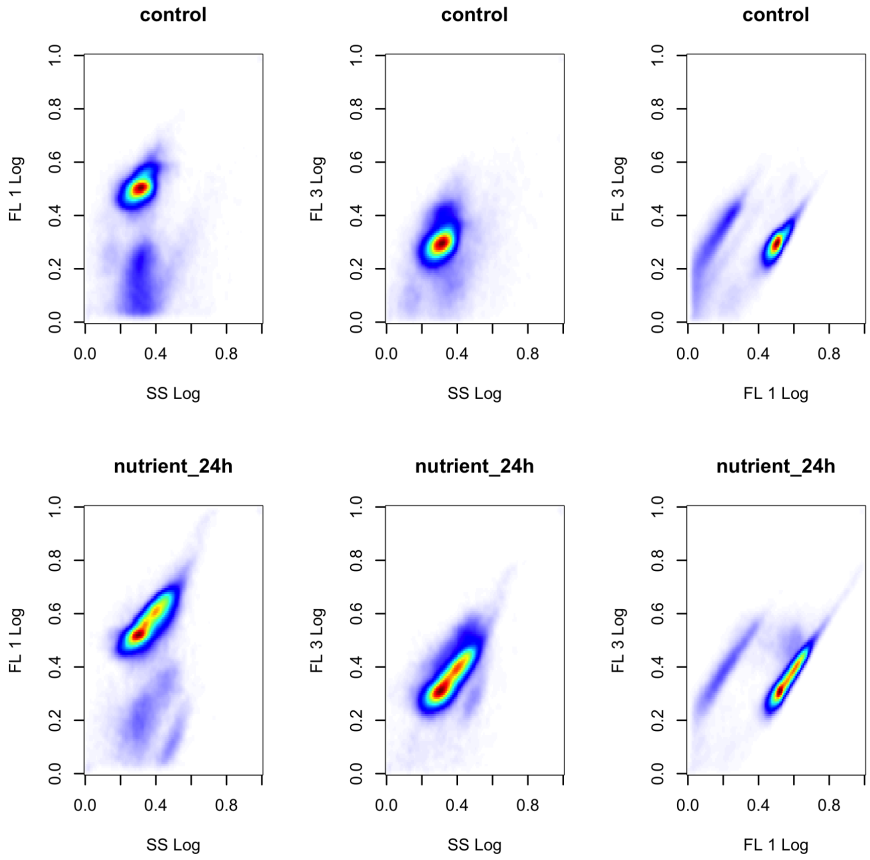

Figure 2: Plot of averaged bivariate density for flows belonging to the control (c) group (top panels) and the 24h nutrient (n24) group (bottom panels). The cells for the c-group are more tightly centered around a SS of 0.35, FL1 of 0.5 and a FL3 of 0.3. For the n24-group the distribution has a larger variance and includes a considerable number of cells at higher SS, FL1 and FL3 values. This region corresponds to larger cells with higher nucleic acids content and intact membranes.

The average contrast of the fingerprints belonging to the 24h nutrient treatment group (n24) and control group (c) is visualised in Figure 3. A negative contrast is represented by light-to-dark blue colours and a positive contrast is indicated with yellow-orange-red colours. After n24-treatment, a lower density of cells is observed at SS of .35, FL1 of .5 and a FL3 of 0.3 as compared to the control treatment, i.e. blue region with a negative contrast. A part of the mass of the distribution shifted to higher SS, FL1 and FL3 values, i.e. yellow-orange region with positive contrast. Contours for negative (blue, less cells) and positive (red, more cells) contrasts can be constructed. They will be useful for the interpretation in the downstream analysis. The contour levels at which contours are drawn are given in the `contourLevel` argument. Here, contours are drawn at -0.04 and 0.04.

```
> par(mfrow=c(2,2))
> L=rep(0,length(group))
> L[group==groupLevels[1]]=-1/sum(group==groupLevels[1])
> L[group==groupLevels[4]]=1/sum(group==groupLevels[4])
> plot(fbasis,L=L,ask=FALSE,main=paste(groupLevels[4],"-",groupLevels[1],sep=""),
+ contour=TRUE,contourLwd=4,contourLevel=c(-.04,.04))
```

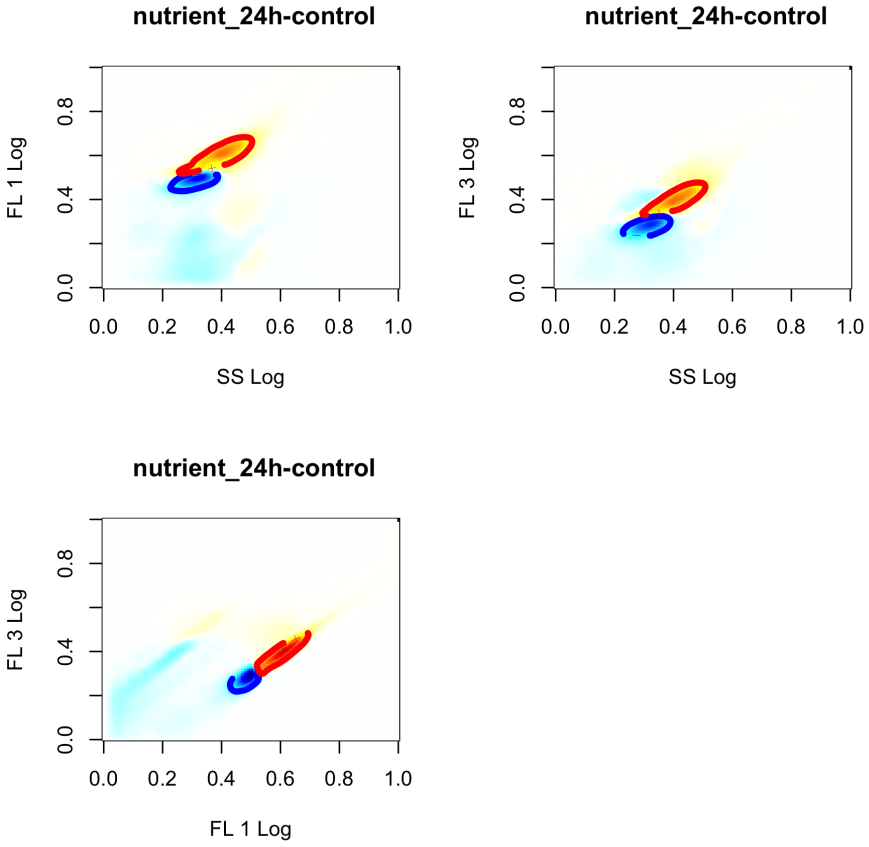

Figure 3: Plot of contrast between kernel density fingerprints belonging to the 24h nutrient (n24) group and the control (c) group. After n24 treatment, a lower density of cells is observed at SS of .35, FL1 of .5 and a FL3 of 0.3, blue region with a negative contrast, as compared to the c-group. A part of the mass of the distribution shifted to higher SS, FL1 and FL3 values, yellow-orange-red region. Contours for negative and positive contrasts can be constructed. They will be useful for the interpretation of the downstream analysis.

The fingerprints can also be explored by using principal component analysis and model based clustering. Principal component analysis (PCA) is often used for dimension reduction. The fingerprint matrix $\boldsymbol{X}$ can be fed into the standard multivariate principal components analysis routine `prcomp`. It performs a rotation of the centered input variables without loss of information and essentially provides a decomposition of the variance-covariance matrix of $\boldsymbol{X}$. We establish dimension reduction by retaining the p principal components (PC's) so that portion of the variance that is explained by them exceeds a certain threshold δ. As PC's consist of linear combinations of input variables, the PC loadings $\boldsymbol{M}$ and scores $\boldsymbol{S}$ can be interpreted in terms of the original pairwise bivariate FC densities. Hence, we can explore the FC data using a low dimensional representation, while still enabling interpretation with respect to features measured in the different FC channels.

For model based clustering we build upon the `mclust` package that adopts Gaussian mixture models. Model based clustering in this example is performed by using the first 6 principal components. They capture more than 95% of the variability in the original fingerprints. Scores on the first 2 principal components are given in Figure 4.

```
> #construct flowPca object
> fPca=flowPca(fbasis)
> #perform model based clustering,
> #use n PCs so as to capture at least 95 % of the variability
> nPca(fPca)<-.95
> nPca(fPca) #number of PCs used for model based clustering
```

```
[1] 6
```

```
> setClust(fPca)<-Mclust(getPcaScore(fPca,nPca(fPca))) #Model based clustering
> cbind(as.character(getClustClass(fPca)),as.character(group)) # cluster class labels and real group
```

```
      [,1] [,2]
 [1,] "1"  "control"
 [2,] "1"  "control"
 [3,] "1"  "control"
 [4,] "1"  "control"
 [5,] "2"  "control"
 [6,] "1"  "control"
 [7,] "2"  "heat_24h"
 [8,] "1"  "heat_24h"
 [9,] "1"  "heat_24h"
[10,] "1"  "heat_24h"
[11,] "1"  "heat_24h"
[12,] "1"  "heat_24h"
[13,] "3"  "heat_3h"
[14,] "3"  "heat_3h"
[15,] "4"  "heat_3h"
[16,] "3"  "heat_3h"
[17,] "4"  "heat_3h"
[18,] "3"  "heat_3h"
[19,] "4"  "nutrient_24h"
[20,] "5"  "nutrient_24h"
[21,] "5"  "nutrient_24h"
[22,] "5"  "nutrient_24h"
```

```
[23,] "5"  "nutrient_24h"
[24,] "5"  "nutrient_24h"
[25,] "6"  "nutrient_3h"
[26,] "6"  "nutrient_3h"
[27,] "6"  "nutrient_3h"
[28,] "6"  "nutrient_3h"
[29,] "6"  "nutrient_3h"
[30,] "6"  "nutrient_3h"
```

```
> par(mfrow=c(1,2))
> plot(fPca,groups=getClustClass(fPca),main="Kernel Dens. (Clustering)")
> plot(fPca,groups=group,main="Kernel Dens. (Treatment)")
> legend("topleft",legend=c("c","h24","h3","n24","n3"),pch=1:5,col=1:5)
```

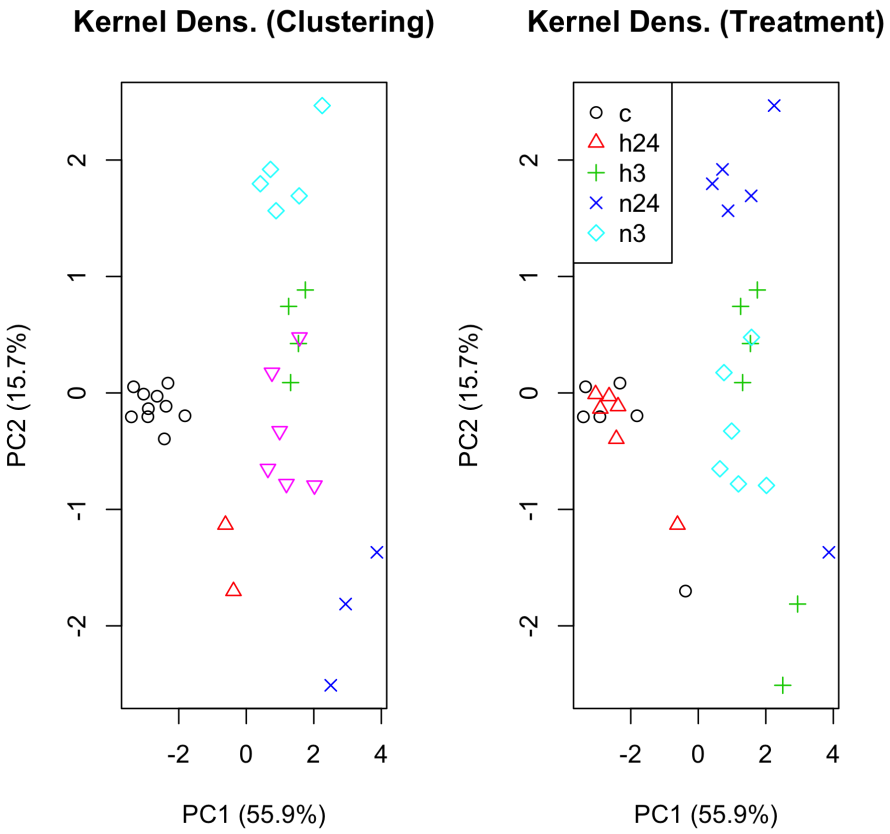

Figure 4: Scores for the first two principal components. In the left panel the samples are classified using model based clustering and in the right panel the samples are labelled according to the actual experimental factor.

The classification with model based clustering provides a considerable resemblance to the real grouping, indicating that main features in the distribution of the size and staining characteristics are affected by the treatment. The separation between 24h heat treatment (h24) and the control (c) treatment in the PC1 and PC2 space indicates that the distributions for the c- and h24-treatment are more similar than the distributions for the flows of other treatments.

The scores $\boldsymbol{S}$ on the PC's can be interpreted in terms of the original bivariate distributions. They consist of linear combinations of the centered fingerprint, the contrast between the sample and average fingerprint over all samples.

$$\boldsymbol{S} = \boldsymbol{X}^*\boldsymbol{M},$$

with $\boldsymbol{X}^* = \boldsymbol{X} - \mathbf{1}\bar{\boldsymbol{X}}^T$ the matrix with the centered fingerprint and $\boldsymbol{M}$ the loading matrix. Some of the $j = 1, \ldots, r$ centered fingerprint grid points x^*_{ij} of sample i will contribute negatively to the k^{th} PC score, s_{ik}, and others positively. The colours in the plot indicates this contribution: $s_{ijk} = x^*_{ij}m_{jk}$. Hence, the score s_{ik} on the k^{th} component consists of the sum over all $n_{\text{bin}} \times n_{\text{bin}}$ grid points in all q bivariate combinations of the flow channels of interest, i.e. $s_{ik} = \sum_{j=1}^{r} x^*_{ij}m_{jk}$. Contours are also added to the interpretation plot. They indicate appropriate contrasts in the original fingerprints.

The interpretation of the score for sample 3 is given in the caption of Figure 5. Because we are assessing one sample, the contours represent the contrast between the sample 3 and the average fingerprint, i.e. the centered fingerprint for sample 3. The contours show if the bivariate kernel density estimate on the grid point is denser (red, "+") or less dens (blue, "-") than the averaged bivariate density estimates over all samples.

Similar plots can be constructed for contrasts in the PCA space. Then the colours will represent the contribution to the contrast on a particular PC and the contours will be constructed for contrasts in centered fingerprints. More details on contrasts are provided in the section on discriminant analysis.

```
> intSamples=3 #for the group average of first group set intSamples=which(group=groupLevels[1])
> par(mfrow=c(2,2))
> plot(fPca,groups=group,main="Treatment")
> pcX=mean(getPca(fPca)$x[intSamples,1])
> pcY=mean(getPca(fPca)$x[intSamples,2])
> arrows(x0=pcX,x1=pcX,y0=-2,y1=pcY)
>  #PCA is done after centering
>  # interpretation in terms of contrast to average bivariate density
>  #contrast between average bivariate density of intSamples vs overall average
> L=rep(-1/nSamp,nSamp)
> L[intSamples]=L[intSamples]+1/length(intSamples)
> plot(fPca,fBasis=fbasis,disc=1,plotType="pcaCont",L=L,ask=FALSE,main="PC 1",contour=TRUE,contourLw
```

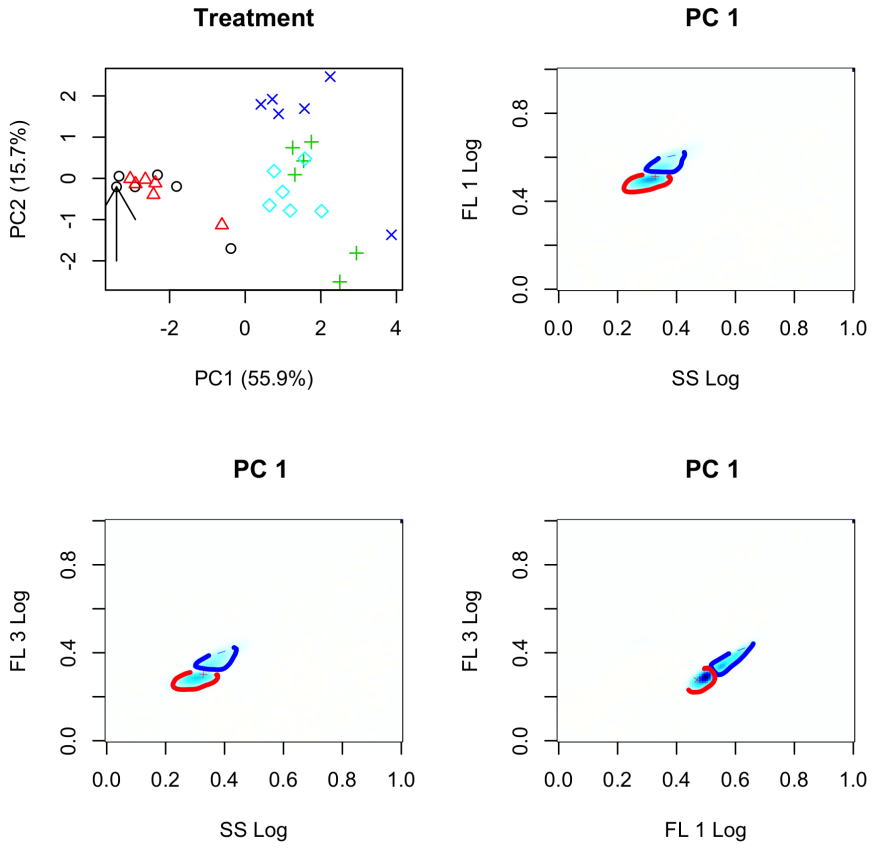

Figure 5: Scores on first two principal components along with an interpretation plot for the score of sample 3 on PC1 (indicated with arrow). The colour in the interpretation plot indicates negative (blue), positive (red) or small (uncoloured) contributions in regions of the bivariate distributions. Overall the colour is neutral or blue, indicating that the score on the first PC will be negative. For a better interpretation, contrast contours are added to the plot. They indicate regions for which the distribution of sample 3 is lower (blue "-" contour)/higher (red "+" contour) than the average bivariate density over all samples. From the interpretation plots it is clear that the FL1 and FL3 bivariate density has the highest contribution to the PC1 score (more intense colouring). The low score on the first PC originates from a density that is higher than average around an SS of 0.3, FL1 of 0.5 and a FL3 of 0.3 (blue colouring within red contour) and because sample 3 has a lower density at higher SS, FL1 and FL3 (blue colour in blue contour), i.e. the low score on the first PC is due to the tightly centered distribution of sample 3 around an SS of 0.35, FL1 of 0.5 and a FL3 of 0.3

5 Discriminant Analysis and Classification

Discriminant analysis (DA) aims at understanding how $K-$groups differ from one another in terms of the $p-$dimensional FC fingerprint. We adopt Fisher DA, which does not impose distributional assumptions and interprets the difference among the $K-$groups by projecting the input variables onto discriminants. In one dimension a good discrimination is obtained when the between class variance of the K experimental groups is large compared to their within class variance. The discriminants are the linear combinations of input variables that maximize the between class variance with respect to the within class variance after projecting the data onto the particular discriminant. It can be shown that K groups in the $r-$dimensional fingerprint feature space span at most a $K-1$ dimensional subspace of orthogonal discriminants. Hence, a huge dimension reduction occurs when $K << r$. Quantitative measures exist for the relative potential of the k-th discriminant function to discriminate the $K-$groups. They are often used for deciding the number of discriminants that are required for a good discrimination between groups and can provide a further dimension reduction. Because discriminants are linear combinations of the input variables, they often provide a very useful interpretation within the original data space. We will use this property for displaying the leading differences in the pairwise bivariate distributions of the FC profiles. If Fisher discriminant analysis (DA) is performed on the original basis, a perfect discrimination will be obtained because we have much more features than observations. Hence, the optimal solution is located in the null space. We provide regularisation of the within group covariance matrix by performing PCA on the fingerprint first and adopting DA on the first p PC's. Note, that PCA is commonly used for regularizing matrix inverses, e.g. in principal component regression to deal with multicollinearity in a linear modelling context, and, that the use of all PC's in the DA would provide the same solution as the DA on all fingerprint features.

We suggest to use the first few PC's that explain more that 95% of the variability in the fingerprint. This can be done by setting `nPca=0.95` when calling the `flowDa` constructor. In our application this corresponds to 6 PCs. Hence, we reduced the dimensionality of the problem from $3 \times 128 \times 128$ fingerprint features to 6.

Since we perform PCA first, the `flowDa` object inherits all properties of a `flowPca` object and all plots from the previous section can be constructed using a `flowDa` object.

```
> #supervised, class labels are needed
> #select first few PC's which explain more than 95% of the variability in the original fingerprint.
> #####Discriminant analysis for kernel dens.
> fDa=flowDa(fbasis,groups= group, nPca=.95)
> fDa


flowDa object
Kernel Density Estimation on grid of 128 x 128

channels
     [,1]       [,2]
[1,] "SS Log"   "FL 1 Log"
[2,] "SS Log"   "FL 3 Log"
[3,] "FL 1 Log" "FL 3 Log"
```

```
> par(mfrow=c(1,2))
> plot(fDa,groups=group,main="Kernel Dens. PCA",plotType="pcaPlot")
> plot(fDa,main="Kernel Dens. DA")
> legend("bottomleft",legend=c("c","h24","h3","n24","n3"),pch=1:5,col=1:5)
```

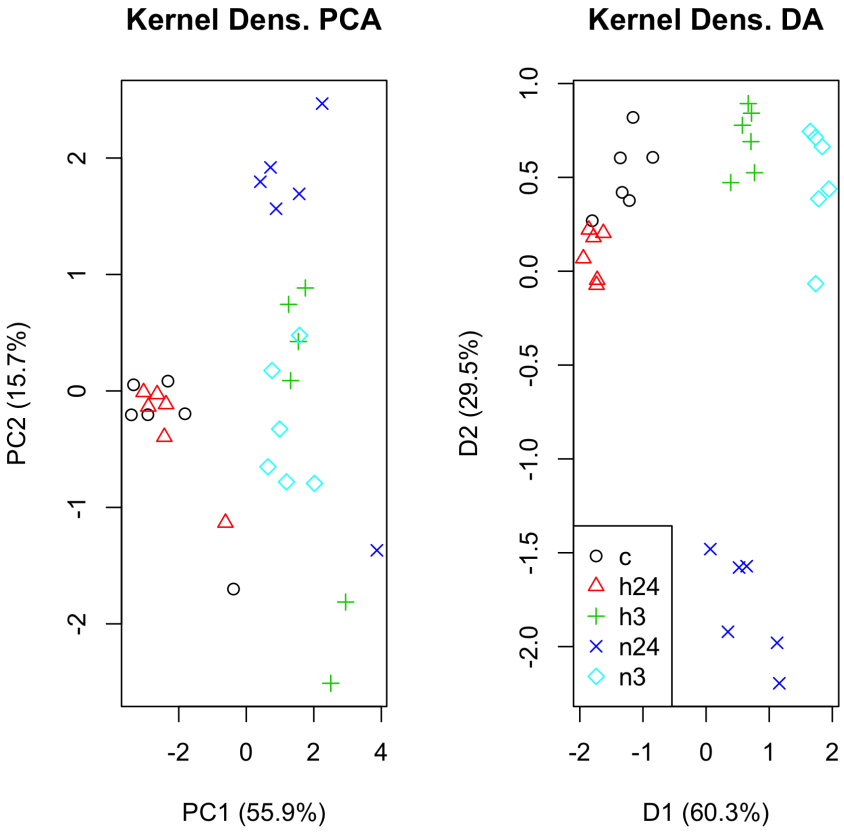

Figure 6: Comparison between clustering of treatments in the PCA space (unsupervised) and in the DA space (supervised).

6 Test for differences in the discriminant space

Statistical hypothesis tests can be performed after dimension reduction. Pairwise tests in the discriminant space are adopted for assessing significance of observed differences between the groups. Standard procedures implemented in `p.adjust()` can be used to address the problem of multiple hypothesis testing. Because the data was already used within the dimension reduction procedure, standard statistical hypothesis testing in the discriminant space does not control the type I error at the nominal significance level, $\alpha = 0.05$. We adopt permutation-based procedures for deriving the null distribution of the test statistics. The following procedure is proposed:

1. Permute class labels,
2. Adopt the PCA-FDA dimension reduction procedure to the permuted data,
3. Construct the permutation-based pairwise test statistics t^* within the discriminant space of 2,
4. Repeat steps 1–3 B times.

Permutation-based p-values are then defined as the fraction of the permuted test statistics that are more extreme than the observed test statistic:

$$p = \frac{\#[|t^*| > |t|]}{B}$$

.

Permutation tests are performed on each discriminant separately so as to retain the interpretation feature: i.e. significant tests can be interpreted in terms of the original fingerprint features.

```
> nPerm=100
> #Only 100 permutations are used
> #so as to restrict the computational burden when generating the vignette
> disc=1:2 #Test only in the space of first 2 discriminants
> fDa=flowDaTest(fDa,disc=disc,nPerm)
```

```
progress
```

Instead of performing more permutations for obtaining accurate p-values, we load the fDa data object supplied in the package to reduce the computational burden. We will adjust the p-values for multiple testing (10 x ndisc tests). Note, that `pAdjustMx` is a wrapper function to the `p.adjust` function, hence with the `method` argument different multiple testing procedures can be adopted. The default of pAdjustMx is the Holm procedure.

```
> data(fDa)
> adjustedPvalues=pAdjustMx(getMpc(fDa)$pValuePerm)
> adjustedPvalues
```

```
                          D1     D2
heat_24h-control      0.4905 0.0110
heat_3h-control       0.0077 0.9380
```

```
nutrient_24h-control     0.0135 0.0000
nutrient_3h-control      0.0013 1.0000
heat_3h-heat_24h         0.0000 0.0000
nutrient_24h-heat_24h    0.0144 0.0000
nutrient_3h-heat_24h     0.0000 0.2862
nutrient_24h-heat_3h     1.0000 0.0013
nutrient_3h-heat_3h      0.0000 0.9380
nutrient_3h-nutrient_24h 0.1071 0.0000
```

Because DA and PCA are both linear projections, significant differences can be interpreted in the original space. Let $\boldsymbol{U}$ denote the $p \times U$ loading matrix of the discriminants, U the number of discriminants, and $\boldsymbol{M}$ the loading matrix of the first p PCs and $\boldsymbol{X}^*$ the centered fingerprint matrix, then an $N \times D$ matrix $\boldsymbol{D}$ with DA scores can be calculated by

$$\boldsymbol{D} = \boldsymbol{X}^* \boldsymbol{M} \boldsymbol{U},$$

and the contrasts $\boldsymbol{C}$ in the DA space by

$$\boldsymbol{C} = \boldsymbol{L}^T \boldsymbol{X}^* \boldsymbol{M} \boldsymbol{U},$$

with $\boldsymbol{L}$ the $N \times L$ contrast matrix and L the number of contrasts of interest. Hence, the contributions, c_{ljk}, to the l^{th} contrast on the k^{th} discriminant D_k at the j^{th} grid point are given by:

$$c_{ljk} = \boldsymbol{L}_l^T \boldsymbol{X}_j^* \boldsymbol{M}_j^T \boldsymbol{U}_k,$$

with $\boldsymbol{L}_l$ the l^{th} column of the contrast matrix $\boldsymbol{L}$, $\boldsymbol{X}_j^*$ the j^{th} column of the centered fingerprint matrix $\boldsymbol{X}^*$, $\boldsymbol{M}_j^T$ the j^{th} row of the PCA loading matrix and $\boldsymbol{U}_k$ the loadings for the k^{th} discriminant. Note, that the l^{th} contrast on the k^{th} discriminant, c_{lk} consists of the sum of the contributions on all r grid points of the fingerprint, $c_{lk} = \sum_{j=1}^{r} c_{ljk}$. By plotting the c_{ljk}, we can interpret the contributions in terms of the original bivariate estimates on the grid points.

This is illustrated in Figure 7 for the contrast between n24 and c-samples on the first discriminant. Note, that in this case the contrast on centered fingerprints and on the original fingerprints will be equal because the average fingerprint used for centering will cancel out. The interpretation is given in the caption of the plot.

```
> nSamp=nSet(fDa)
> L<-rep(0,nSamp)
> L[group==groupLevels[4]]<-1/sum(group==groupLevels[4])
> L[group==groupLevels[1]]<--1/sum(group==groupLevels[1])
> par(mfrow=c(2,2))
> plot(fDa)
> legend("bottomleft",legend=c("c","h24","h3","n24","n3"),pch=1:5,col=1:5)
> disc=1
> plot(fDa,fBasis=fbasis,L=L,ask=FALSE,plotType="discCont",disc=disc,
+ contour=TRUE,contourLevel=c(-.04,.04),contourLwd=4)
```

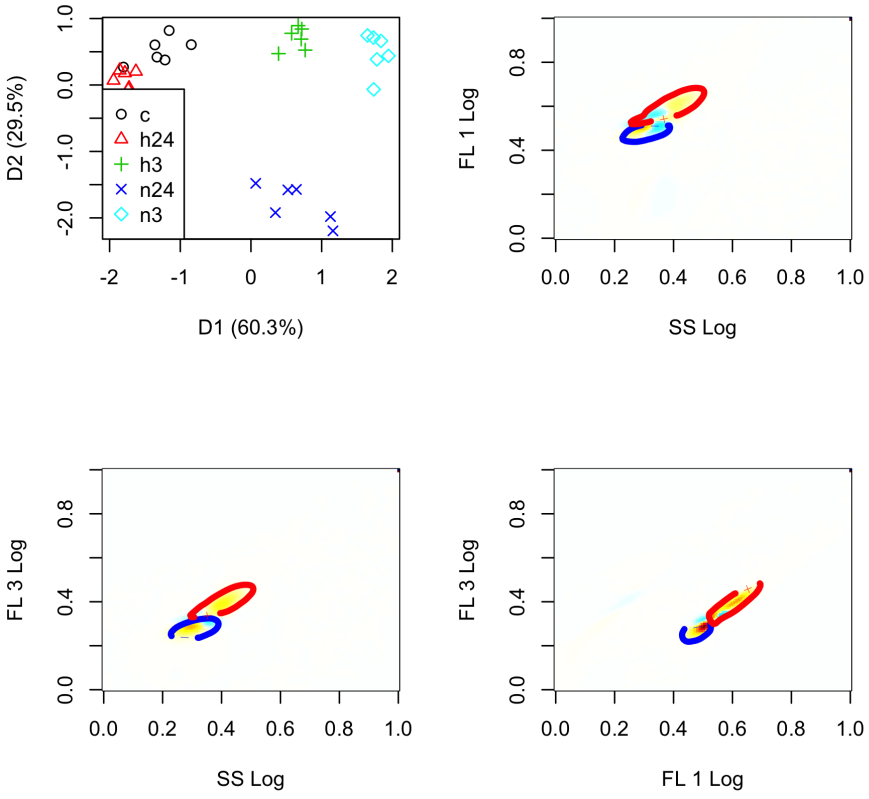

Figure 7: Interpretation of 24h nutrient and control (n24-c) contrast on the first discriminant (D1). The colours in the interpretation plots indicate the contribution of the region to the n24-c contrast on D1. Negative contributions are represented with a light-to-dark blue colour scheme, positive contributions are coloured in yellow-orange-red. The contours on the plots indicate regions for which the contrasts in the fingerprints is negative (less cells, blue contour) or positive (more cells, red contour). In the plot it can be seen that n24-samples on average score higher on D1 than c-samples. The difference in D1 score is thus linked to a lower abundance of cells in the n24 condition around SS of 0.35, FL1 of 0.5 and a FL3 of 0.3 and a higher abundance at higher SS, FL1 and FL3 values than that of the average fingerprint of c-samples.

The significant difference between 24h heat (h24) and control (c) treatment is more subtle. The score on the second discriminant is on average slightly lower for h24 than for c-samples. The interpretation plots show that this corresponds to a shift of the h24 distribution to slightly lower SS and a slightly higher FL1. (Figure 8)

```
> nSamp=nSet(fDa)
> L<-rep(0,nSamp)
> L[group==groupLevels[2]]<-1/sum(group==groupLevels[2])
> L[group==groupLevels[1]]<--1/sum(group==groupLevels[1])
> par(mfrow=c(2,2))
> plot(fDa)
> legend("bottomleft",legend=c("c","h24","h3","n24","n3"),pch=1:5,col=1:5)
> disc=2
> plot(fDa,fBasis=fbasis,L=L,ask=FALSE,plotType="discCont",disc=disc,
+ contour=TRUE,contourLevel=c(-.04,.04),contourLwd=4)
```

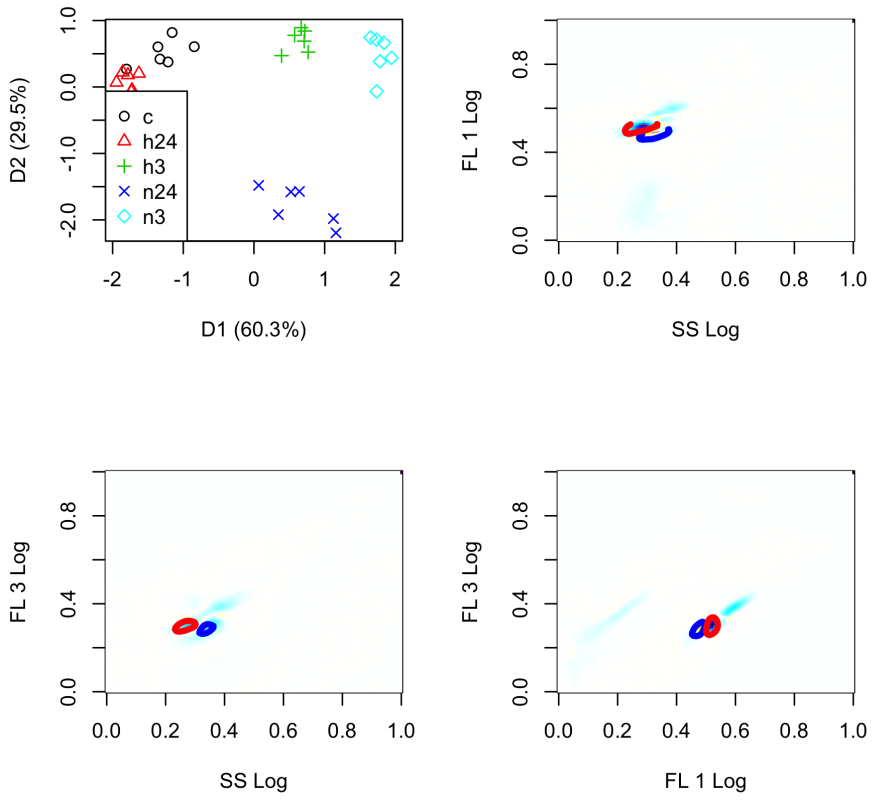

Figure 8: Interpretation of 24h heat and control (h24-c) contrast on the second discriminant (D2). The colours in the interpretation plots indicate the contribution of the region to the h24-c contrast on D2. Negative contributions are represented with a light-to-dark blue colour scheme, positive contributions are coloured in yellow-orange-red. The contours on the plots indicate regions where the contrasts in the fingerprints is negative (less cells, blue contour) or positive (more cells red contour). In the plot it can be seen that h24-samples on average score lower on D2 than c-samples. The difference in D2 score is linked to a subtle shift towards slightly lower SS and slightly higher FL1 values.

7 Generating plots for multiple contrasts

For a good comparison between contrasts, the same colour scheme has to be used. This can be done by fixing the `colorLimit` argument. If you generate plots using a matrix with multiple contrasts, this is done automatically. The contrasts are structured in the columns, i.e. the matrix has the same number of rows as the number of samples and each column represents a different contrast.

```
> library(flowFDAExampleData)
> library(flowFDA)
> data(fbasis)
> data(fDa)
> nSamp=nSet(fDa)
> group=getGroups(fDa)
> nGroup=nlevels(group)
> groupLevels=levels(group)
> sampleNames=rownames(getBasis(fbasis))
> ###############################
> #Generate original Fingerprint plots
> ###############################
>
> #uncomment to plot all bivariate distributions
> #par(mfrow=c(1,3))
> #for (i in 1:nSamp) plot(fbasis,sample=i,ask=TRUE,main=sampleNames[i])
>
> #uncomment to create a pdf with all bivariate distributions plots
> #pdf("allBasis.pdf",height=7,width=15)
> #par(mfrow=c(1,3))
> ##cex to enlarge font
> #for (i in 1:nSamp)
> #plot(fbasis,sample=i,ask=FALSE,main=sampleNames[i],
> #cex.main=1.5,cex.axis=1.5,cex.lab=1.5)
> #dev.off()
>
>
> #uncomment to plot all average bivariate distributions for each group
> #par(mfrow=c(1,3))
> #for (i in 1:nGroup) plot(fbasis,sample=group==groupLevels[i],ask=TRUE,main=groupLevels[i])
>
> #uncomment to create a pdf with all average bivariate distributions plots for each group
> #pdf("allGroupBasis.pdf",height=7,width=15)
> #par(mfrow=c(1,3))
> #for (i in 1:nGroup)
> #plot(fbasis,sample=group==groupLevels[i],ask=FALSE,main=groupLevels[i],
> #cex.main=1.5,cex.axis=1.5,cex.lab=1.5)
> #dev.off()
>
> ###############################################
> #Contrast interpretation plots
> ###############################################
>
> #extract groups involved in contrasts from p value object
```

```
> comp=strsplit(rownames(getMpc(fDa)$p),split="-")
> #create all corresponding contrasts
> L=sapply(comp,function(x,group)
+ (group==x[1])/sum(group==x[1])-(group==x[2])/sum(group==x[2]),group=group)
> colnames(L)<-rownames(getMpc(fDa)$p)
> #uncomment to generate contrast plot of flowBasis Object
> #par(mfrow=c(1,3))
> #plot(fbasis,L=L,ask=TRUE,contour=TRUE,contourLwd=4,contourLevel=c(-.04,.04))
>
> #uncomment to generate a pdf of the contrast plot in original space
> #pdf("contrastInterpretationPlotsBasis.pdf",height=7,width=15)
> #par(mfrow=c(1,3))
> #plot(fbasis,L=L,ask=FALSE,contour=TRUE,contourLwd=4,contourLevel=c(-.04,.04),
> #cex.main=1.5,cex.axis=1.5,cex.lab=1.5)
> #dev.off()
>
> adjustedPvalues=pAdjustMx(getMpc(fDa)$p)
> #uncomment to make discriminant interpretation plots in plot window
> #par(mfrow=c(1,3))
> #disc=1
> #plot(fDa,fBasis=fbasis,L=L,ask=TRUE,plotType="discCont",disc=disc,contour=TRUE,
> #contourLevel=c(-.04,.04),contourLwd=4,
> #main=paste("\n D", disc," p=",round(adjustedPvalues[,disc],3),sep=""))
> #disc=2
> #plot(fDa,fBasis=fbasis,L=L,ask=TRUE,plotType="discCont",disc=disc,contour=TRUE,
> #contourLevel=c(-.04,.04),contourLwd=4,
> #main=paste("\n D", disc," p=",round(adjustedPvalues[,disc],3),sep=""))
>
> #uncomment to create a pdf of the discriminant interpretation plots of contrasts
> #pdf("contrastInterpretationPlotsDiscriminant.pdf",height=7,width=15)
> #par(mfrow=c(1,3))
> #for (disc in 1:2)
> #plot(fDa,fBasis=fbasis,L=L,ask=FALSE,plotType="discCont",disc=disc,contour=TRUE,
> #contourLevel=c(-.04,.04),contourLwd=4,
> #main=paste("\n D", disc," p=",round(getMpc(fDa)$p[,disc],3),sep=""),
> #cex.main=1.5,cex.axis=1.5,cex.lab=1.5)
> #dev.off()
```

References

De Roy, K., Clement, L., Thas, O., Wang, Y., and Boon, N. (2012). Flow cytometry for fast microbial community fingerprinting. Water Research, 46 (3), 907-919.

Ellis, B., Haaland, P., Hahne, F., Le Meur, N. and Gopalakrishnan, N. (2009). flowCore: flowCore: Basic structures for flow cytometry data. R package version 1.26.3.

Ramsay, J. and Silverman, B. (2005). Functional Data Analysis. Second Edition, Springer, New York.