pdfCluster: Partial Discharges Clustering

Oscar Perpiñán Lamigueiro Miguel Angel Sánchez Urán

17 May 2011

```
> library(pdCluster)
```

The set of examples will use dataset which is loaded with:

```
> load("~/Investigacion/PD/Datos/20100922/DescargasRAW.RData")
```

1 Feature generation

1.1 Prony

A clean partial discharge signal can be regarded as a finite combination of damped complex exponentials. Under this assumption, the so-called Prony's method allows for the estimation of frequency, amplitude, phase and damping components of the signal [KT82, HDS90, KTS84].

Let's use one of the signals from the dataset (figure 1). This signal contains zeros at the beginning and at the end. The `no0` function can remove these parts of the signal (figure 2).

```
> signal <- lista[[3]]
```

With this cleaned signal the Prony's method can provide their components (figure 3). Since the number of components must be fixed *a priori*, the function `compProny` allows the comparison of different numbers (figure 4)

1.2 Feature generation

The `pdCluster` includes several functions for feature generation. The `analysis` functions comprises all of them. The results for our example signal are:

```
> analysis(signal)

  RefMax           W1          W2          W3        W4        range   N
1    154 1.585665e-07 2.80655e-05 0.004660942 0.2967786 0.0009498277 323
        energy        nZC   freq1   freq2   damp1   damp2
1 8.301006e-06 0.04643963 3071675 1704751 1140738 3385233
```

This function can be used with a list of signals in order to obtain a matrix of features:

```
> analysisList <- lapply(lista[1:10], analysis)
> pdData <- do.call(rbind, analysisList)
```

Now we need the angle and reflection information, available from another different file. In order to safely share the information, both data frames must be reordered by their energy values:

```
> pdSummary <- read.csv("~/Investigacion/PD/Datos/20100922/descargas.csv")[1:10,
+     ]
> idxOrderSummary = order(pdSummary$sumaCuadrados)
> idxOrderData = order(pdData$energy)
> pdDataOrdered = cbind(pdData[idxOrderData, ], pdSummary[idxOrderSummary,
+     c("angulo", "separacionOriginal")])
```

Later, the data frame to be used with the clustering algorithm has to ordered by time. Thus the samples of the `clara` method will be random.

```
> idx <- do.call(order, pdSummary[idxOrderSummary, c("segundo",
+     "inicio")])
> pdDataOrdered <- pdDataOrdered[idx, ]
```

We can now construct a `PD` object[1].

[1] The `pdCluster` package is designed with S4 classes and methods. Two classes have been defined: `PD` and `PDCluster`.

```
> plot(signal, type = "l")
```

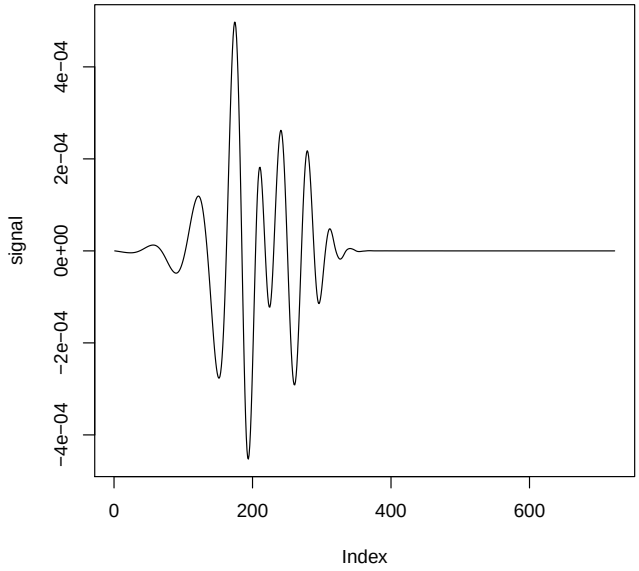

Figure 1: Partial discharge signal

```
> plot(no0(signal), type = "l")
```

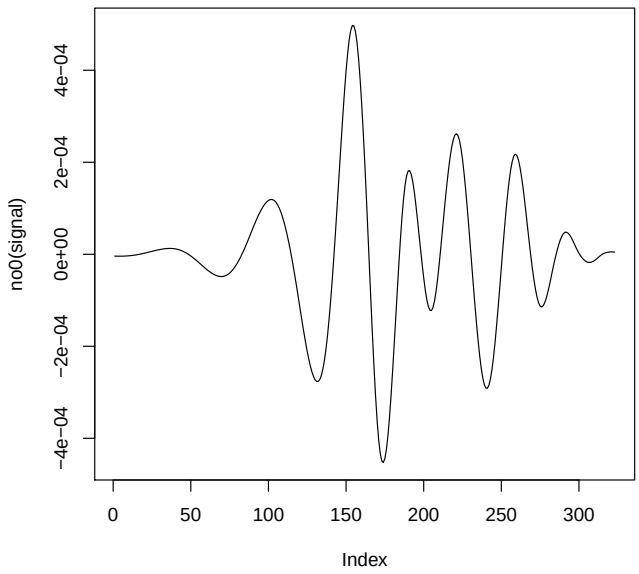

Figure 2: Partial discharge signal after a threshold cleaning

```
> p <- prony(signal, M = 10)
> print(plot(p))
```

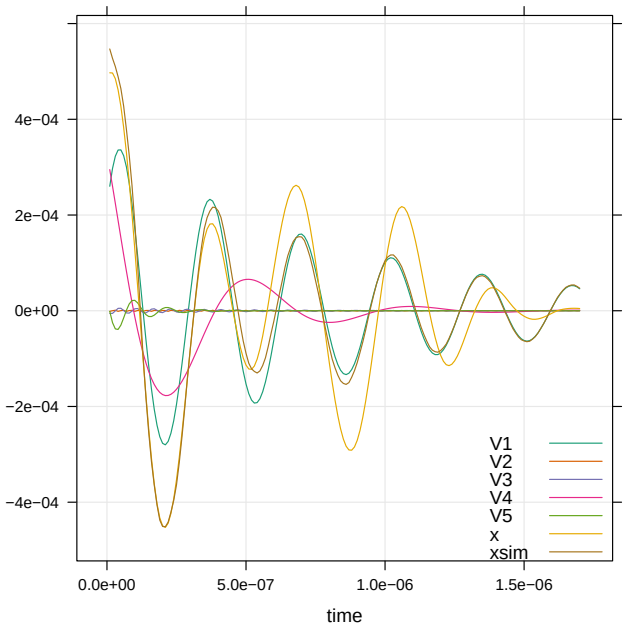

Figure 3: Prony's method results

```
> p <- compProny(signal, M = c(10, 20, 30, 40))
> print(p)
```

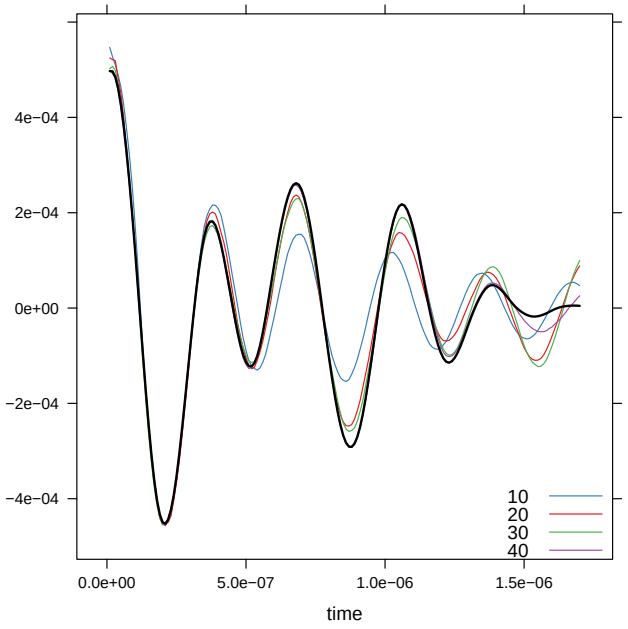

Figure 4: Comparison of different Prony decompositions

```
> pd <- df2PD(pdDataOrdered)
> pd

Object of class  PD

Source of measurements:
Number of observations:  10
Filtered?:  FALSE
Transformed?: FALSE

Data:
     RefMax             W1                  W2                  W3
 Min.   :  10.0   Min.   :8.417e-14   Min.   :1.487e-11   Min.   :2.875e-09
 1st Qu.:  56.5   1st Qu.:3.521e-08   1st Qu.:6.808e-06   1st Qu.:1.158e-03
 Median : 176.0   Median :2.128e-07   Median :2.981e-05   Median :4.565e-03
 Mean   : 320.0   Mean   :2.423e-04   Mean   :6.606e-03   Mean   :3.205e-02
 3rd Qu.: 212.0   3rd Qu.:5.609e-07   3rd Qu.:9.860e-05   3rd Qu.:9.841e-03
 Max.   :1152.0   Max.   :2.234e-03   Max.   :6.465e-02   Max.   :2.749e-01
       W4                range                 N              energy
 Min.   :0.0000000   Min.   :8.272e-10   Min.   :  86.0   Min.   :9.544e-18
 1st Qu.:0.0007306   1st Qu.:2.792e-04   1st Qu.: 230.8   1st Qu.:2.038e-06
 Median :0.2051125   Median :1.428e-03   Median : 328.0   Median :1.208e-05
 Mean   :0.1893751   Mean   :7.392e-03   Mean   : 594.7   Mean   :6.944e-04
 3rd Qu.:0.3037326   3rd Qu.:1.021e-02   3rd Qu.: 374.5   3rd Qu.:6.760e-04
 Max.   :0.5386891   Max.   :3.070e-02   Max.   :1954.0   Max.   :3.282e-03
      nZC               freq1             damp1             angle
 Min.   :0.005249   Min.   : 354724   Min.   : 175077   Min.   :73.40
 1st Qu.:0.020226   1st Qu.:1370271   1st Qu.: 679157   1st Qu.:74.39
 Median :0.030888   Median :2731837   Median :2755132   Median :81.99
 Mean   :0.033111   Mean   :2210311   Mean   :2320361   Mean   :79.44
 3rd Qu.:0.044445   3rd Qu.:3066132   3rd Qu.:3635957   3rd Qu.:82.19
 Max.   :0.070093   Max.   :3113590   Max.   :4715804   Max.   :84.82
    refl
 Mode :logical
 FALSE:4
 TRUE :6
 NA's :0
```

The results of `analysis` to the whole dataset are available with:

```
> load("~/Investigacion/PD/Datos/20100922/dfHibr17112010.RData")
```

```
> dfHibr <- df2PD(dfHibr)
> dfHibr

Object of class  PD

Source of measurements:
Number of observations:  9955
Filtered?:  FALSE
Transformed?: FALSE

Data:
     RefMax              W1                  W2                  W3
 Min.   :    0.0   Min.   :0.000e+00   Min.   :0.000e+00   Min.   :0.000e+00
 1st Qu.:   57.0   1st Qu.:6.838e-14   1st Qu.:1.228e-11   1st Qu.:2.378e-09
 Median :  205.0   Median :8.011e-13   Median :1.250e-10   Median :1.730e-08
 Mean   :  559.4   Mean   :1.197e-03   Mean   :2.366e-03   Mean   :1.352e-02
 3rd Qu.:  896.0   3rd Qu.:1.104e-07   3rd Qu.:1.740e-05   3rd Qu.:2.009e-03
 Max.   :34079.0   Max.   :2.973e-01   Max.   :1.165e-01   Max.   :5.869e-01
       W4                range                 N               energy
 Min.   :0.000e+00   Min.   :0.000e+00   Min.   :    0.0   Min.   :0.000e+00
 1st Qu.:3.341e-07   1st Qu.:2.265e-05   1st Qu.:  172.0   1st Qu.:2.833e-08
 Median :1.103e-06   Median :1.064e-04   Median :  417.0   Median :5.665e-07
 Mean   :7.648e-02   Mean   :5.624e-03   Mean   :  992.1   Mean   :4.023e-02
 3rd Qu.:3.686e-02   3rd Qu.:1.026e-03   3rd Qu.: 1568.0   3rd Qu.:1.307e-05
 Max.   :1.430e+00   Max.   :3.108e-01   Max.   :34234.0   Max.   :5.459e+00
      nZC               freq1              damp1               angle
 Min.   :0.000000   Min.   :       0   Min.   :       0   Min.   :  0.01422
 1st Qu.:0.003629   1st Qu.:  299379   1st Qu.:   178107   1st Qu.: 77.08625
 Median :0.005682   Median :  384840   Median :   427577   Median :227.66600
 Mean   :0.013704   Mean   : 1250969   Mean   :  1807198   Mean   :184.30712
 3rd Qu.:0.021283   3rd Qu.: 1849583   3rd Qu.:  1625622   3rd Qu.:264.08250
 Max.   :0.147059   Max.   :50000000   Max.   :205053894   Max.   :359.93300
    refl
 Mode :logical
 FALSE:4329
 TRUE :5626
 NA's :0
```

2 Transformations

Prior to the clustering algorithm, the feature matrix has to be filtered:

```
> dfFilter <- filterPD(dfHibr)
```

and transformed [BC64]:

```
> dfTrans <- transformPD(dfFilter)
> dfTrans

Object of class  PD

Source of measurements:
Number of observations:  3695
Filtered?:  TRUE
Transformed?: TRUE

Data:
     RefMax             W1                W2                W3
 Min.   : 0.000   Min.   :-53.341   Min.   :-38.720   Min.   :-27.9936
 1st Qu.: 6.809   1st Qu.:-45.027   1st Qu.:-29.386   1st Qu.:-20.6000
 Median : 7.896   Median :-26.477   Median :-18.205   Median :-13.0398
 Mean   : 8.387   Mean   :-30.009   Mean   :-19.236   Mean   :-12.7681
 3rd Qu.:10.407   3rd Qu.:-18.402   3rd Qu.:-10.554   3rd Qu.: -5.1754
 Max.   :18.979   Max.   : -1.496   Max.   : -2.249   Max.   : -0.5339
       W4               range              N              energy
 Min.   :-20.5419   Min.   :-31.952   Min.   :3.381   Min.   :-59.712
 1st Qu.:-14.3313   1st Qu.:-12.589   1st Qu.:3.934   1st Qu.:-19.287
 Median : -7.3826   Median :-10.500   Median :4.171   Median :-16.453
 Mean   : -8.1065   Mean   : -9.551   Mean   :4.217   Mean   :-14.776
 3rd Qu.: -1.2579   3rd Qu.: -6.206   3rd Qu.:4.561   3rd Qu.: -8.933
 Max.   :  0.1203   Max.   : -1.210   Max.   :5.339   Max.   :  1.650
      nZC              freq1           damp1            angle
 Min.   :-12.686   Min.   :4.996   Min.   : 6.581   Min.   :  0.01422
 1st Qu.: -9.468   1st Qu.:5.131   1st Qu.:10.800   1st Qu.: 74.98095
 Median : -8.107   Median :5.226   Median :11.327   Median :202.19300
 Mean   : -7.424   Mean   :5.216   Mean   :11.527   Mean   :178.20110
 3rd Qu.: -5.080   3rd Qu.:5.313   3rd Qu.:12.434   3rd Qu.:264.47700
 Max.   : -3.268   Max.   :5.433   Max.   :14.677   Max.   :359.93300
    refl
 Mode :logical
 FALSE:3695
 NA's :0
```

The figure 5 compares the datasets after and before of the transformations:

```
> nZCbefore <- as.data.frame(dfFilter)$nZC
> nZCafter <- as.data.frame(dfTrans)$nZC
> comp <- data.frame(After = nZCafter, Before = nZCbefore)
```

2.1 Graphical tools

The `pdCluster` packages includes a set of graphical exploratory tools, such as a scatterplot matrices with hexagonal binning [CLNL87] (figure 6), density plots (figure 7), histograms (figure 8) or phase resolved partial discharge patterns, both with partial transparency (figure 9) or hexagonal binning (figure 10).

3 Clustering

The filtered and transformed object can now be used with the clustering algorithm [SHR97]. The results are displayed with a phase resolved pattern with clusters in separate clusters in the figure 11, and with superposed clusters in the figure 12.

The results can be easily understood with the density plots of each cluster and feature (figure 13) or with the histograms (figure 14).

```
> h <- histogram(~After + Before, data = comp, scales = list(x = list(relation = "free"),
+     y = list(relation = "free", draw = FALSE)), breaks = 100,
+     col = "gray", xlab = "", strip.names = c(TRUE, TRUE), bg = "gray",
+     fg = "darkblue")
> print(h)
```

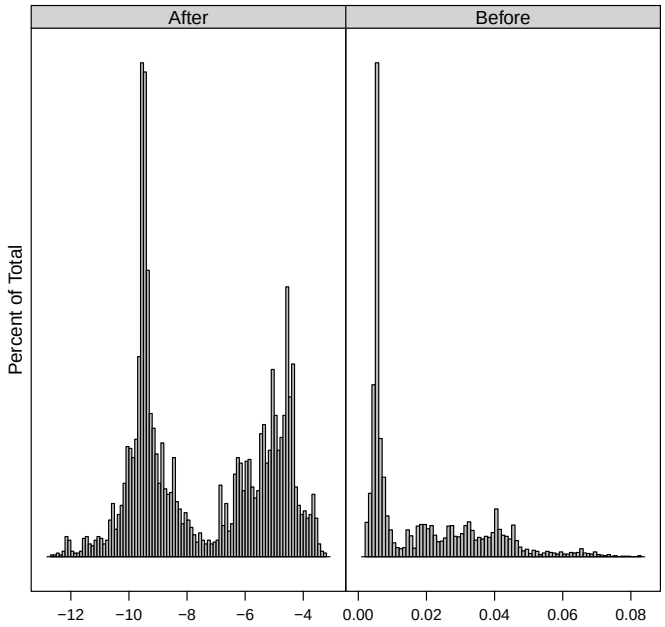

Figure 5: Histogram of a collection of partial discharges

```
> p <- splom(dfTrans)
> print(p)
```

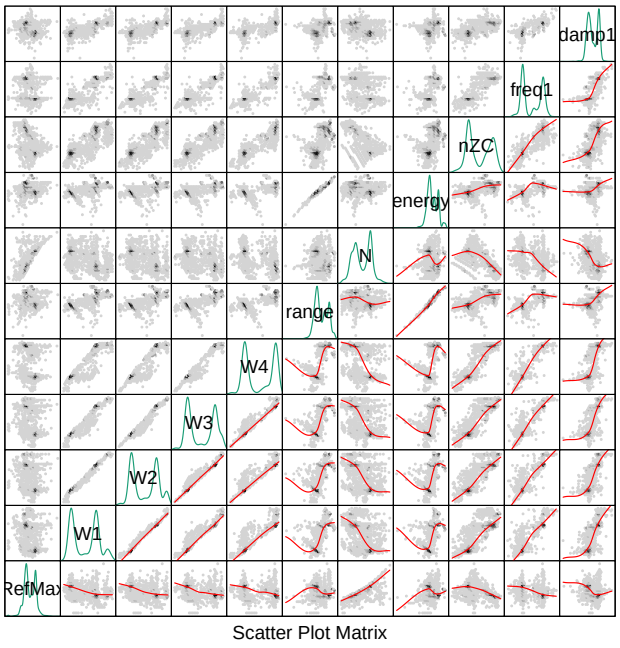

Figure 6: Scatterplot matrix of a collection of partial discharges

```
> p <- densityplot(dfTrans)
> print(p)
```

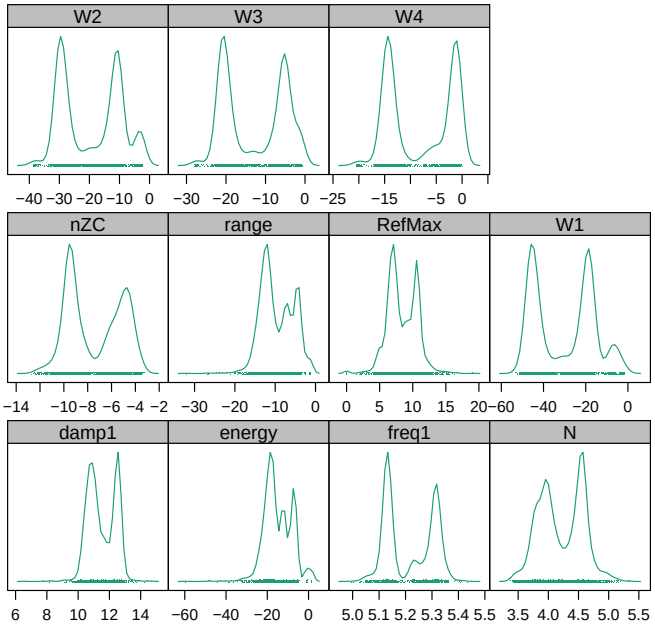

Figure 7: Density plot of a collection of partial discharges

```
> p <- histogram(dfTrans)
> print(p)
```

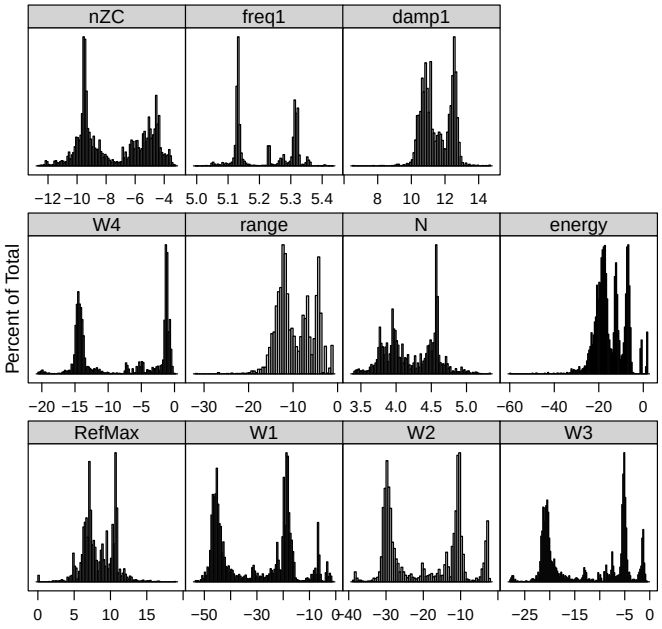

Figure 8: Histograms of a collection of partial discharges

```
> p <- xyplot(dfTrans)
> print(p)
```

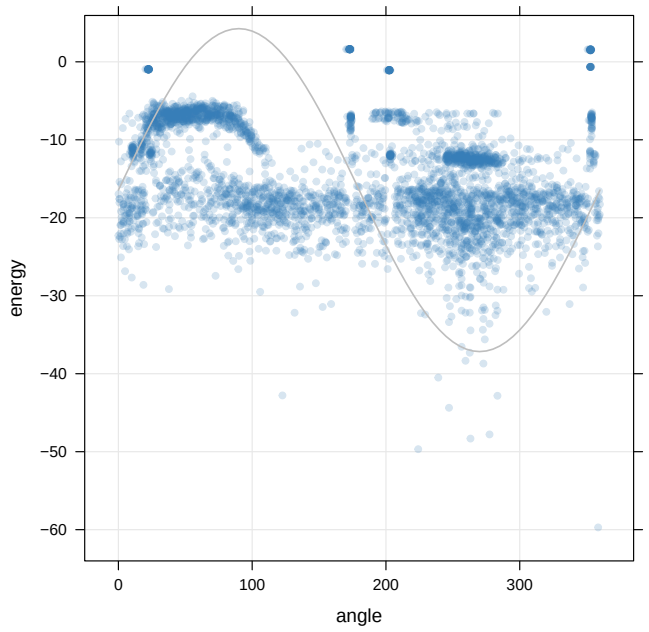

Figure 9: Partial discharge phase resolved pattern.

```
> p <- hexbinplot(dfTrans)
> print(p)
```

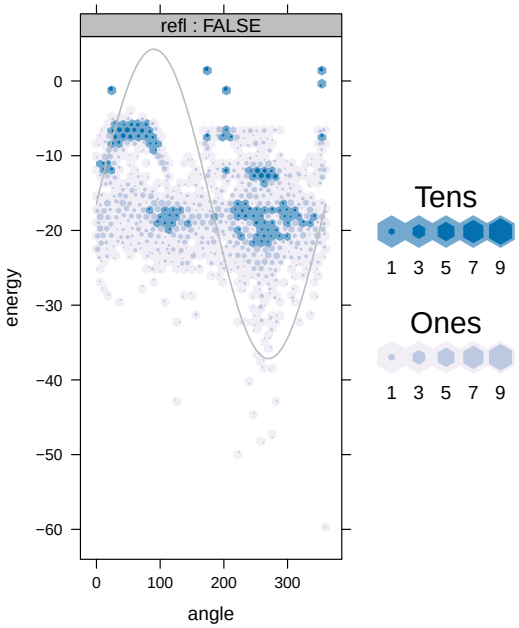

Figure 10: Partial discharge phase resolved pattern with hexbinplot

```
> dfTransCluster <- claraPD(dfTrans, noise.level = 0.7, noise.rm = TRUE)
> dfTransCluster

Object of class  PD

Source of measurements:
Number of observations:  2644
Filtered?:  TRUE
Transformed?: TRUE

Data:
     RefMax             W1                W2                W3
 Min.   : 1.972   Min.   :-53.341   Min.   :-38.720   Min.   :-27.994
 1st Qu.: 7.081   1st Qu.:-45.515   1st Qu.:-29.703   1st Qu.:-20.887
 Median : 9.025   Median :-35.130   Median :-24.943   Median :-19.048
 Mean   : 8.842   Mean   :-31.586   Mean   :-20.195   Mean   :-13.380
 3rd Qu.:10.612   3rd Qu.:-18.439   3rd Qu.:-10.554   3rd Qu.: -5.171
 Max.   :16.753   Max.   : -6.286   Max.   : -2.715   Max.   : -1.162
       W4                range               N             energy
 Min.   :-20.54193   Min.   :-20.107   Min.   :3.381   Min.   :-32.977
 1st Qu.:-14.48296   1st Qu.:-12.493   1st Qu.:3.943   1st Qu.:-19.014
 Median :-13.66034   Median :-10.977   Median :4.422   Median :-16.648
 Mean   : -8.43826   Mean   : -9.292   Mean   :4.273   Mean   :-14.153
 3rd Qu.: -1.23238   3rd Qu.: -4.736   3rd Qu.:4.569   3rd Qu.: -7.572
 Max.   : -0.09173   Max.   : -1.210   Max.   :5.202   Max.   :  1.650
      nZC              freq1           damp1            angle
 Min.   :-12.192   Min.   :5.005   Min.   : 6.581   Min.   :  0.7043
 1st Qu.: -9.523   1st Qu.:5.131   1st Qu.:10.706   1st Qu.: 64.0166
 Median : -8.819   Median :5.139   Median :11.135   Median :171.9235
 Mean   : -7.414   Mean   :5.211   Mean   :11.460   Mean   :166.1924
 3rd Qu.: -4.954   3rd Qu.:5.312   3rd Qu.:12.475   3rd Qu.:259.4360
 Max.   : -3.268   Max.   :5.329   Max.   :13.115   Max.   :359.9330
    refl
 Mode :logical
 FALSE:2644
 NA's :0



Number of clusters:  7
Number of elements per cluster:

   1    2    3    4
1384  696  420  144

Metric:  manhattan
Number of simulations:  25
Noise level:  0.7
Distances
      dist            distRel         distFactor
 Min.   : 0.1707   Min.   :0.0000   Min.   :  1.000
 1st Qu.: 2.1393   1st Qu.:0.7161   1st Qu.:  3.000
 Median : 4.8503   Median :0.8677   Median :  5.000
 Mean   : 6.8810   Mean   :0.8098   Mean   :  4.984
 3rd Qu.:11.1163   3rd Qu.:0.9414   3rd Qu.:  7.000
 Max.   :50.0681   Max.   :0.9956   Max.   :  9.000
                                    NA's   :265.000
```

References

[BC64] G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):pp. 211–252, 1964.

[CLNL87] D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield. Scatterplot matrix techniques for large n. *Journal of the American Statistical Association*, 82(398):pp. 424–436, 1987.

[HDS90] J.F. Hauer, C.J. Demeure, and L.L. Scharf. Initial results in prony analysis of power system response signals. *Power Systems, IEEE Transactions on*, 5(1):80 –89, feb 1990.

[KT82] R. Kumaresan and D. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 30(6):833 – 840, dec. 1982.

[KTS84] R. Kumaresan, D.W. Tufts, and L.L. Scharf. A prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models. *Proceedings of the IEEE*, 72(2):230 – 233, feb. 1984.

[SHR97] Anja Struyf, Mia Hubert, and Peter Rousseeuw. Clustering in an object-oriented environment. *Journal of Statistical Software*, 1(4):1–30, 2 1997.

```
> p <- xyplot(dfTransCluster)
> print(p)
```

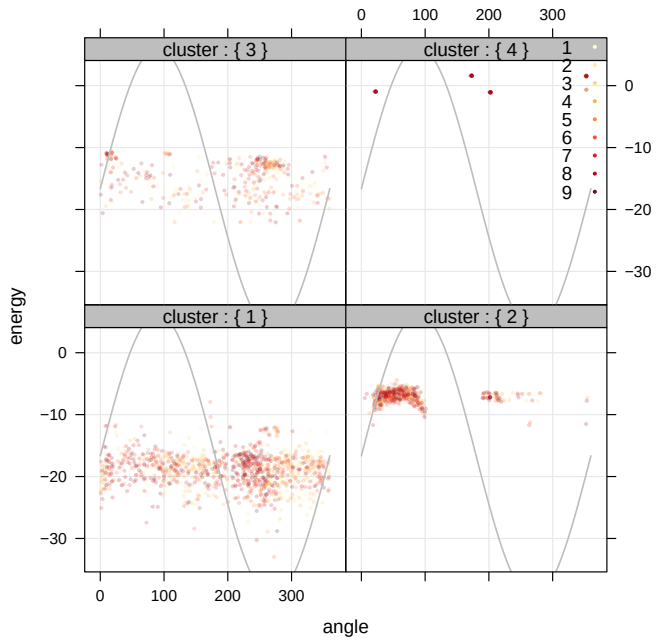

Figure 11: Partial discharge phase resolved pattern with clusters in separate panels.

```
> p <- xyplot(dfTransCluster, panelClust = FALSE)
> print(p)
```

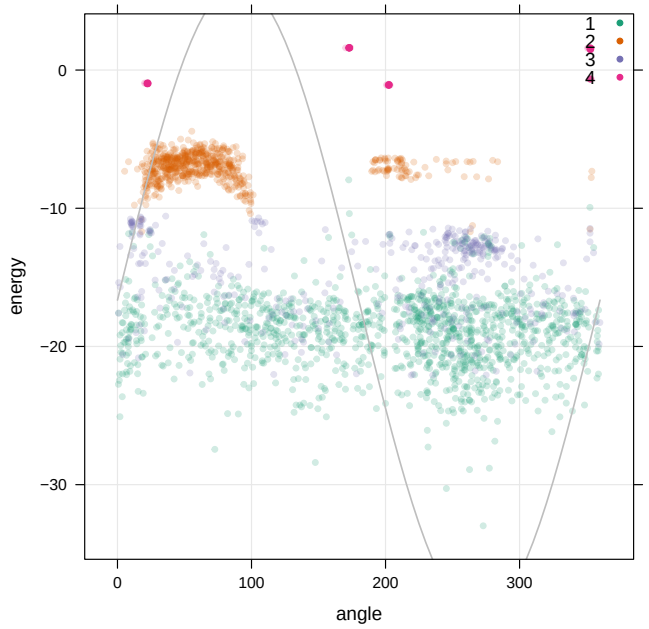

Figure 12: Partial discharge phase resolved pattern with clusters marked with colors.

```
> p <- densityplot(dfTransCluster)
> print(p)
```

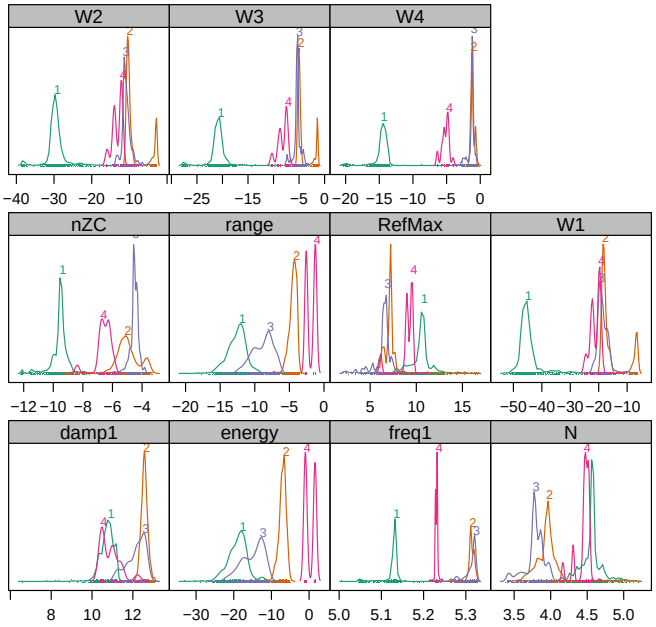

Figure 13: Density plot of the clusters of partial discharges

```
> p <- histogram(dfTransCluster)
> print(p)
```

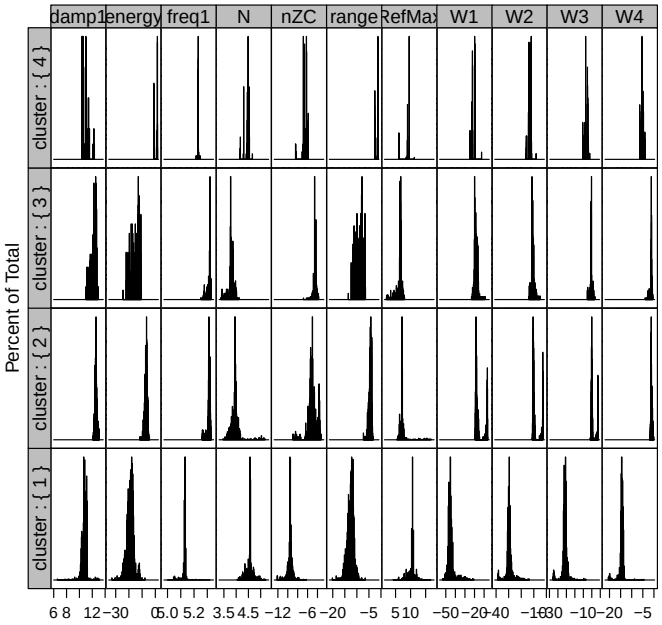

Figure 14: Histograms of the clusters of partial discharges