Transformation and other Factors of the pairwise Mass Spectrometry peak-list Comparison Process.(PRELIMINARY!!!)

Peter Martus and Witold Eryk Wolski

April 10, 2006

Anova analysis of the PMF dataset

We analyse here the 4 datasets (pmf.binary ,pmf.intensity, msms.binary, msms.intensity). The datasets provide results of evaluating the sensitivity and specificity of the pairwise peak-list comparison performed on an dataset of identified Tandem MS data (msms) and on an dataset of identified Peptide Mass Fingerprint spectra. (publication submitted).

Load the results for the binary measures.

```
> rm(list = ls())
> library(msbase)
> data("pmf.binary")
> range(pmf.binary$TPPAUC)

[1] 75.51390 97.92409

> range(pmf.binary$FPPAUC)

[1] 34.33935 99.62211
```

The minimal linear model (containing as few factors as possible) which sufficiently describes the outcome the specificity-PAUC (given small FP rates) of the experiment is given by.

Specificity PAUC – FACTORS

```
> tplm <- lm(FPPAUC ~ Measure + Theta + Length + Weight + Noncross,
+     data = pmf.binary)
> summary(tplm)$adj.r

[1] 0.3173803

> anova(tplm)

Analysis of Variance Table

Response: FPPAUC
```

```
          Df    Sum Sq   Mean Sq   F value    Pr(>F)
Measure    3    2712.3     904.1    4.7187 0.0042434 **
Theta      2    4621.2    2310.6   12.0595 2.384e-05 ***
Length     1    2662.1    2662.1   13.8942 0.0003434 ***
Weight     1       0.1       0.1    0.0004 0.9843654
Noncross   1 2.253e-04 2.253e-04 1.176e-06 0.9991373
Residuals 87   16669.1     191.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Measure    3.2   10.2    14.9      NA     NA
Theta      2.1   17.3    38.1      NA     NA
Length     1.1   10.0    43.9      NA     NA
Weight     1.1    0.0     0.0      NA     NA
Noncross   1.1    0.0     0.0      NA     NA
Residuals 91.6   62.5     3.2      NA     NA
```

Sensitivity PAUC – FACTORS

```
> tplm <- lm(TPPAUC ~ Measure + Theta + Length + Weight + Noncross,
+     data = pmf.binary)
> summary(tplm)$adj.r
```

```
[1] 0.3189317
```

```
> anova(tplm)
```

```
Analysis of Variance Table

Response: TPPAUC
          Df    Sum Sq   Mean Sq   F value    Pr(>F)
Measure    3    305.34    101.78    5.0618 0.0028040 **
Theta      2    477.67    238.83   11.8780 2.749e-05 ***
Length     1    272.29    272.29   13.5417 0.0004038 ***
Weight     1      0.07      0.07    0.0037 0.9516113
Noncross   1 1.788e-05 1.788e-05 8.894e-07 0.9992497
Residuals 87   1749.33     20.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Measure    3.2   10.9    16.1      NA     NA
Theta      2.1   17.0    37.7      NA     NA
```

```
Length      1.1    9.7    43.0       NA     NA
Weight      1.1    0.0     0.0       NA     NA
Noncross    1.1    0.0     0.0       NA     NA
Residuals 91.6   62.4     3.2       NA     NA
```

Specificity PAUC – FINAL

```
> tplm <- lm(FPPAUC ~ Measure + Theta + Length + Measure * Theta +
+     Measure * Length + Theta * Length + Measure * Theta * Length,
+     data = pmf.binary)
> summary(tplm)$adj.r

[1] 0.9999614

> anova(tplm)

Analysis of Variance Table

Response: FPPAUC
                     Df Sum Sq Mean Sq F value    Pr(>F)
Measure               3 2712.3   904.1   83343 < 2.2e-16 ***
Theta                 2 4621.2  2310.6  212999 < 2.2e-16 ***
Length                1 2662.1  2662.1  245403 < 2.2e-16 ***
Measure:Theta         6 4675.3   779.2   71831 < 2.2e-16 ***
Measure:Length        3 2697.4   899.1   82884 < 2.2e-16 ***
Theta:Length          2 4621.9  2311.0  213032 < 2.2e-16 ***
Measure:Theta:Length  6 4673.9   779.0   71809 < 2.2e-16 ***
Residuals            72    0.8 0.01085
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> tmp1 <- round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)[,
+     2:3]
```

Sensitivity PAUC – FINAL

```
> tplm <- lm(TPPAUC ~ Measure + Theta + Length + Measure * Theta +
+     Measure * Length + Theta * Length + Measure * Theta * Length,
+     data = pmf.binary)
> summary(tplm)$adj.r

[1] 0.9996686

> anova(tplm)

Analysis of Variance Table

Response: TPPAUC
                     Df Sum Sq Mean Sq F value    Pr(>F)
Measure               3 305.34  101.78 10401.2 < 2.2e-16 ***
```

```
Theta                 2 477.67  238.83 24407.4 < 2.2e-16 ***
Length                1 272.29  272.29 27826.1 < 2.2e-16 ***
Measure:Theta         6 495.25   82.54  8435.3 < 2.2e-16 ***
Measure:Length        3 280.64   93.55  9560.1 < 2.2e-16 ***
Theta:Length          2 477.54  238.77 24400.9 < 2.2e-16 ***
Measure:Theta:Length  6 495.27   82.54  8435.6 < 2.2e-16 ***
Residuals            72   0.70    0.01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
                       Df Sum Sq Mean Sq F value Pr(>F)
Measure               3.2   10.9     9.2      NA     NA
Theta                 2.1   17.0    21.5      NA     NA
Length                1.1    9.7    24.5      NA     NA
Measure:Theta         6.3   17.7     7.4      NA     NA
Measure:Length        3.2   10.0     8.4      NA     NA
Theta:Length          2.1   17.0    21.5      NA     NA
Measure:Theta:Length  6.3   17.7     7.4      NA     NA
Residuals            75.8    0.0     0.0      NA     NA
```

```
> tmp2 <- cbind(round(sweep(as.matrix(dr), 2, ddd, "/") * 100,
+     digits = 1)[, 2:3])
```

To identify the best measure we compute the average Sensitivity-PAUC for each CP having the same length , theta and measure factor.

```
> with(pmf.binary, tapply(TPPAUC, list(Length = Length, Theta = Theta,
+     Measure = Measure), mean))
```

```
, , Measure = fm

      Theta
Length      0.5        1        2
   0   97.91767 97.91652 97.91795
   250 97.91767 97.91652 97.91795

, , Measure = gower

      Theta
Length      0.5        1        2
   0   97.86121 97.86196 97.86121
   250 97.86121 97.86196 97.86121

, , Measure = hg

      Theta
Length      0.5        1        2
```

```
   0   97.68596 97.70260 97.69902
   250 97.89561 96.66816 75.81042

, , Measure = rmi

      Theta
Length      0.5        1        2
   0   97.81934 97.81609 97.80941
   250 97.84059 96.87455 81.02380
```

The average given this three factors is computed for the Specificitiy-PAUC.

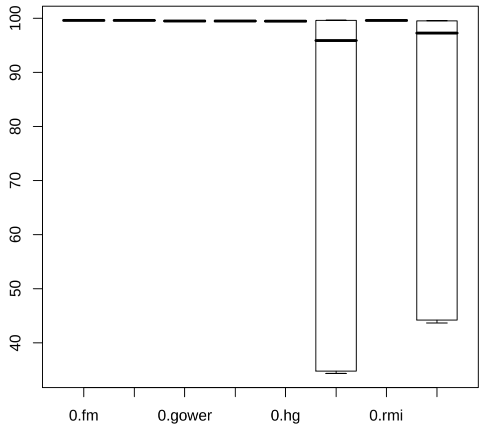
```
> with(pmf.binary, tapply(FPPAUC, list(Length = Length, Theta = Theta,
+     Measure = Measure), mean))
```

```
, , Measure = fm

      Theta
Length      0.5        1        2
   0   99.60086 99.60786 99.61083
   250 99.60086 99.60786 99.61083

, , Measure = gower

      Theta
Length      0.5      1      2
   0   99.49289 99.493 99.4929
   250 99.49289 99.493 99.4929

, , Measure = hg

      Theta
Length      0.5        1        2
   0   99.46382 99.46557 99.46289
   250 99.61564 95.88389 34.56005

, , Measure = rmi

      Theta
Length      0.5        1        2
   0   99.60031 99.59965 99.59956
   250 99.53658 97.26311 43.94947
```

Looking at the output of the tapply function we identified the Fowlkes Mallows statistics as the best measure.

Study the interaction between length and measure.

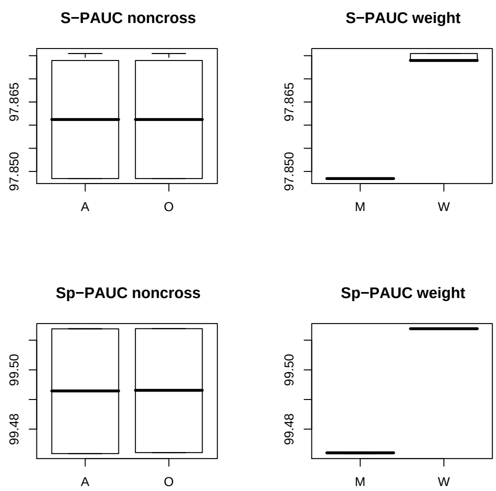
```
> boxplot(FPPAUC ~ Length * Measure, data = pmf.binary)
> par(mar = c(6, 3, 3, 3))
> boxplot(TPPAUC ~ Length * Theta * Measure, data = pmf.binary,
+     las = 2, ylim = c(95, 100))
> abline(v = 1:50, col = "gray")
```



Fowlkes mallows

Take a look on which other factors the performance of the fm measure depends.

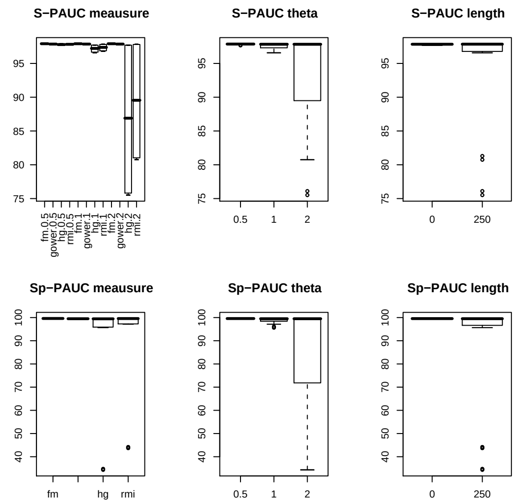
```
> bingow <- pmf.binary[pmf.binary$Measure == "gower", ]
> par(mfrow = c(2, 2))
> boxplot(TPPAUC ~ Noncross, data = bingow, main = "S-PAUC noncross")
> boxplot(TPPAUC ~ Weight, data = bingow, main = "S-PAUC weight")
> boxplot(FPPAUC ~ Noncross, data = bingow, main = "Sp-PAUC noncross")
> boxplot(FPPAUC ~ Weight, data = bingow, main = "Sp-PAUC weight")
```



Make some more boxplots on what the outcome of the comparison depends.

```
> par(mfcol = c(2, 3))
> boxplot(TPPAUC ~ Measure * Theta, data = pmf.binary, main = "S-PAUC meausure",
```

```
+     las = 2)
> boxplot(FPPAUC ~ Measure, data = pmf.binary, main = "Sp-PAUC meausure")
> boxplot(TPPAUC ~ Theta, data = pmf.binary, main = "S-PAUC theta")
> boxplot(FPPAUC ~ Theta, data = pmf.binary, main = "Sp-PAUC theta")
> boxplot(TPPAUC ~ Length, data = pmf.binary, main = "S-PAUC length")
> boxplot(FPPAUC ~ Length, data = pmf.binary, main = "Sp-PAUC length")
```



The figure makes it clear that the measure rmi and hg strongly depend on the choice of the theta.

PMF data - intensity based maesures

Load the evaluation results.

```
> data("pmf.intensity")
```

The minimal model explaining as much as possible variance is:
Specificity PAUC

```
> intlm <- lm(FPPAUC ~ Measure + Scale + Theta + Length + Weight +
+     Noncross + Trans, data = pmf.intensity)
> anova(intlm)
```

```
Analysis of Variance Table

Response: FPPAUC
            Df    Sum Sq   Mean Sq   F value    Pr(>F)
Measure      6    657114    109519  203.1894 < 2.2e-16 ***
Scale        3    410669    136890  253.9694 < 2.2e-16 ***
Theta        2     80009     40005   74.2203 < 2.2e-16 ***
Length       1     12295     12295   22.8106 1.884e-06 ***
Weight       1       452       452    0.8385    0.3599
Noncross     1 5.281e-03 5.281e-03 9.797e-06    0.9975
```

```
Trans         3     11507      3836    7.1165 9.259e-05 ***
Residuals 2670   1439130       539
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> apply(anova(intlm), 2, sum)
```

```
       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 2611176.2  303534.9        NA        NA
```

```
> summary(intlm)$adj.r
```

```
[1] 0.4453485
```

```
> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Measure    0.2   25.2    36.1      NA     NA
Scale      0.1   15.7    45.1      NA     NA
Theta      0.1    3.1    13.2      NA     NA
Length     0.0    0.5     4.1      NA     NA
Weight     0.0    0.0     0.1      NA     NA
Noncross   0.0    0.0     0.0      NA     NA
Trans      0.1    0.4     1.3      NA     NA
Residuals 99.4   55.1     0.2      NA     NA
```

Sensitivity PAUC

```
> intlm <- lm(TPPAUC ~ Measure + Scale + Theta + Weight + Noncross +
+     Length + Trans, data = pmf.intensity)
> anova(intlm)
```

```
Analysis of Variance Table

Response: TPPAUC
            Df    Sum Sq   Mean Sq   F value    Pr(>F)
Measure      6    269703     44951  157.2068 < 2.2e-16 ***
Scale        3    300341    100114  350.1302 < 2.2e-16 ***
Theta        2      8793      4396   15.3755 2.294e-07 ***
Weight       1        40        40    0.1383    0.7100
Noncross     1 7.620e-04 7.620e-04 2.665e-06    0.9987
Length       1      4964      4964   17.3605 3.190e-05 ***
Trans        3       895       298    1.0431    0.3723
Residuals 2670    763440       286
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> apply(anova(intlm), 2, sum)
```

```
       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 1348175.5  155048.2        NA        NA
```

```
> summary(intlm)$adj.r

[1] 0.4301179

> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

            Df Sum Sq Mean Sq F value Pr(>F)
Measure    0.2   20.0    29.0      NA     NA
Scale      0.1   22.3    64.6      NA     NA
Theta      0.1    0.7     2.8      NA     NA
Weight     0.0    0.0     0.0      NA     NA
Noncross   0.0    0.0     0.0      NA     NA
Length     0.0    0.4     3.2      NA     NA
Trans      0.1    0.1     0.2      NA     NA
Residuals 99.4   56.6     0.2      NA     NA
```

Specificity PAUC

```
> intlm <- lm(FPPAUC ~ Measure + Scale + Theta + Length + Measure:Scale +
+     Measure:Theta + Measure:Length, data = pmf.intensity)
> anova(intlm)

Analysis of Variance Table

Response: FPPAUC
                 Df Sum Sq Mean Sq F value    Pr(>F)
Measure           6 657114  109519 789.070 < 2.2e-16 ***
Scale             3 410669  136890 986.270 < 2.2e-16 ***
Theta             2  80009   40005 288.228 < 2.2e-16 ***
Length            1  12295   12295  88.583 < 2.2e-16 ***
Measure:Scale    18 873389   48522 349.591 < 2.2e-16 ***
Measure:Theta    12 164012   13668  98.474 < 2.2e-16 ***
Measure:Length    6  47408    7901  56.928 < 2.2e-16 ***
Residuals      2639 366280     139
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> apply(anova(intlm), 2, sum)

       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 2611176.2  368937.5        NA        NA

> summary(intlm)$adj.r

[1] 0.8571745

> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
                Df Sum Sq Mean Sq F value Pr(>F)
Measure        0.2   25.2    29.7      NA     NA
Scale          0.1   15.7    37.1      NA     NA
Theta          0.1    3.1    10.8      NA     NA
Length         0.0    0.5     3.3      NA     NA
Measure:Scale  0.7   33.4    13.2      NA     NA
Measure:Theta  0.4    6.3     3.7      NA     NA
Measure:Length 0.2    1.8     2.1      NA     NA
Residuals     98.2   14.0     0.0      NA     NA
```

Sensitivity PAUC

```
> intlm <- lm(TPPAUC ~ Measure + Scale + Theta + Length + Measure:Scale +
+     Measure:Theta + Measure:Length, data = pmf.intensity)
> anova(intlm)
```

```
Analysis of Variance Table

Response: TPPAUC
                 Df Sum Sq Mean Sq  F value    Pr(>F)
Measure           6 269703   44951  881.545 < 2.2e-16 ***
Scale             3 300341  100114 1963.373 < 2.2e-16 ***
Theta             2   8793    4396   86.219 < 2.2e-16 ***
Length            1   4964    4964   97.350 < 2.2e-16 ***
Measure:Scale    18 555255   30848  604.964 < 2.2e-16 ***
Measure:Theta    12  26239    2187   42.883 < 2.2e-16 ***
Measure:Length    6  48316    8053  157.923 < 2.2e-16 ***
Residuals      2639 134564      51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> apply(anova(intlm), 2, sum)
```

```
       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 1348175.5  195562.3        NA        NA
```

```
> summary(intlm)$adj.r
```

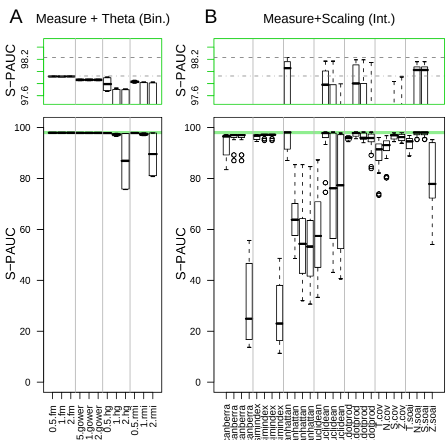
```
[1] 0.8983724
```

```
> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
                Df Sum Sq Mean Sq F value Pr(>F)
Measure        0.2   20.0    23.0      NA     NA
Scale          0.1   22.3    51.2      NA     NA
Theta          0.1    0.7     2.2      NA     NA
Length         0.0    0.4     2.5      NA     NA
Measure:Scale  0.7   41.2    15.8      NA     NA
Measure:Theta  0.4    1.9     1.1      NA     NA
Measure:Length 0.2    3.6     4.1      NA     NA
Residuals     98.2   10.0     0.0      NA     NA
```

Now comes the code to generate a quite complex image.

```
> nf <- layout(matrix(c(1, 2, 3, 4), 2, 2, byrow = TRUE), c(1.5,
+     2.4), c(1, 3), TRUE)
> par(mar = c(1, 3.5, 3, 1))
> boxplot(TPPAUC ~ Theta + Measure, data = pmf.binary, las = 2,
+     axes = F, ylim = c(97.5, 98.5), cex.axis = 0.9)
> axis(2, cex.axis = 0.9, col = 3)
> box(, col = 3)
> abline(v = c(3.5, 6.5, 9.5), col = "gray70")
> abline(h = max(pmf.intensity$TPPAUC), lty = 2, col = "gray50")
> abline(h = max(pmf.binary$TPPAUC), lty = 4, col = "gray50")
> mtext("S-PAUC", side = 2, line = 2, cex = 1)
> mtext("Measure + Theta (Bin.)", side = 3, line = 1)
> mtext("A", side = 3, line = 1, at = -3, cex = 1.5)
> par(mar = c(1, 3, 3, 1))
> boxplot(TPPAUC ~ Scale + Measure, data = pmf.intensity, las = 2,
+     axes = F, ylim = c(97.5, 98.5), cex.axis = 0.9)
> abline(v = c(4.5, 8.5, 12.5, 16.5, 20.5, 24.5), col = "gray70")
> abline(v = max(pmf.binary$TPPAUC))
> axis(2, cex.axis = 0.9, col = 3)
> box(, col = 3)
> abline(h = max(pmf.intensity$TPPAUC), lty = 2, col = "gray50")
> abline(h = max(pmf.binary$TPPAUC), lty = 4, col = "gray50")
> mtext("S-PAUC", side = 2, line = 2, cex = 1)
> mtext("Measure+Scaling (Int.)", side = 3, line = 1)
> mtext("B", side = 3, line = 1, at = -1, cex = 1.5)
> par(mar = c(6, 3.5, 0, 1))
> plot(1, 1, xlim = c(0, 13), ylim = c(0, 100), type = "n", axes = F,
+     ylab = "", xlab = "", main = "")
> rect(-1, 97.5, 14, 98.5, col = "lightgreen", border = "lightgreen")
> par(new = TRUE)
> boxplot(TPPAUC ~ Theta + Measure, data = pmf.binary, las = 2,
+     ylim = c(0, 100), ylab = "", cex.axis = 0.9)
> mtext("S-PAUC", side = 2, line = 2)
> abline(v = c(3.5, 6.5, 9.5), col = "gray70")
> par(mar = c(6, 3, 0, 1))
> plot(1, 1, xlim = c(0, 25), ylim = c(0, 100), type = "n", axes = F,
+     ylab = "", xlab = "", main = "")
> rect(-4, 97.5, 30, 98.5, col = "lightgreen", border = "lightgreen")
> par(new = TRUE)
> boxplot(TPPAUC ~ Scale + Measure, data = pmf.intensity, las = 2,
+     type = "n", ylim = c(0, 100), cex.axis = 0.9)
> mtext("S-PAUC", side = 2, line = 2, cex = 1)
> abline(v = c(4.5, 8.5, 12.5, 16.5, 20.5, 24.5), col = "gray70")
```



Now we tabulate the scores for the Specificity-PAUC according to the identified factors.

```
> with(pmf.intensity, tapply(FPPAUC, list(Length = Length, Theta = Theta,
+     Scale = Scale, Measure = Measure), mean, data = pmf.intensity))

, , Scale = T, Measure = canberra

      Theta
Length      0.5       1        2
   0    40.91063 93.6597 94.85207
   250  40.91063 93.6597 94.85207

, , Scale = N, Measure = canberra

      Theta
Length      0.5        1       2
   0    75.47692 94.57132 94.9742
   250  75.47692 94.57132 94.9742

, , Scale = S, Measure = canberra

      Theta
Length      0.5        1       2
   0    75.47692 94.57132 94.9742
   250  75.47692 94.57132 94.9742

, , Scale = Z, Measure = canberra

      Theta
Length       0.5         1         2
   0    14.838601 16.970899 18.025241
   250   6.668776  6.875478  6.987514

, , Scale = T, Measure = simindex
```

```
      Theta
Length      0.5        1        2
   0   84.18635 94.16826 94.87792
   250 84.18635 94.16826 94.87792

, , Scale = N, Measure = simindex

      Theta
Length      0.5        1        2
   0   92.47554 94.81671 95.00968
   250 92.47554 94.81671 95.00968

, , Scale = S, Measure = simindex

      Theta
Length      0.5        1        2
   0   92.47554 94.81671 95.00968
   250 92.47554 94.81671 95.00968

, , Scale = Z, Measure = simindex

      Theta
Length      0.5         1         2
   0   11.60160 12.200991 12.577445
   250  5.54774  5.559005  5.568968

, , Scale = T, Measure = manhattan

      Theta
Length      0.5        1        2
   0   38.37919 99.66254 99.70945
   250 38.37919 99.66254 99.70945

, , Scale = N, Measure = manhattan

      Theta
Length     0.5        1        2
   0   30.3108 50.15505 60.21486
   250 30.3108 50.15505 60.21486

, , Scale = S, Measure = manhattan

      Theta
Length      0.5        1        2
   0   20.43354 30.02840 35.75854
   250 30.31080 50.15505 60.21486

, , Scale = Z, Measure = manhattan
```

```
      Theta
Length      0.5        1        2
   0    16.20285 25.29194 31.08887
   250 25.30977 43.37358 53.49222

, , Scale = T, Measure = euclidean

      Theta
Length      0.5        1        2
   0    5.436158 36.56946 53.94575
   250 5.436158 36.56946 53.94575

, , Scale = N, Measure = euclidean

      Theta
Length      0.5        1        2
   0    64.17502 99.40814 99.22895
   250 64.17502 99.40814 99.22895

, , Scale = S, Measure = euclidean

      Theta
Length      0.5        1        2
   0    25.35236 47.83727 54.55622
   250 64.17502 99.40814 99.22895

, , Scale = Z, Measure = euclidean

      Theta
Length      0.5        1        2
   0    21.28464 43.82615 50.80906
   250 60.68594 98.56850 98.61628

, , Scale = T, Measure = dotprod

      Theta
Length      0.5        1        2
   0    98.95265 98.95265 98.95265
   250 98.95265 98.95265 98.95265

, , Scale = N, Measure = dotprod

      Theta
Length      0.5        1        2
   0    99.40779 99.40779 99.40779
   250 99.40779 99.40779 99.40779

, , Scale = S, Measure = dotprod

      Theta
```

```
Length      0.5        1        2
   0   97.29015 97.29015 97.29015
   250 99.40779 99.40779 99.40779

, , Scale = Z, Measure = dotprod

      Theta
Length      0.5        1        2
   0   96.76746 93.03602 86.67268
   250 99.41862 99.35492 98.58796

, , Scale = T, Measure = cov

      Theta
Length      0.5        1        2
   0   82.27304 68.76517 57.05578
   250 82.27304 68.76517 57.05578

, , Scale = N, Measure = cov

      Theta
Length      0.5        1        2
   0   88.48299 78.89884 68.17747
   250 88.48299 78.89884 68.17747

, , Scale = S, Measure = cov

      Theta
Length     0.5        1        2
   0   97.7699 97.25855 96.21299
   250 97.7699 97.25855 96.21299

, , Scale = Z, Measure = cov

      Theta
Length      0.5        1        2
   0   96.68121 95.15267 93.11853
   250 96.68121 95.15267 93.11853

, , Scale = T, Measure = soai

      Theta
Length     0.5        1        2
   0   83.4165 85.04398 85.94792
   250 83.4165 85.04398 85.94792

, , Scale = N, Measure = soai

      Theta
Length      0.5       1       2
```

```
  0   97.58838 97.7956 97.9013
  250 97.58838 97.7956 97.9013

, , Scale = S, Measure = soai

      Theta
Length      0.5       1       2
  0   97.58838 97.7956 97.9013
  250 97.58838 97.7956 97.9013

, , Scale = Z, Measure = soai

      Theta
Length      0.5        1        2
  0   72.98227 69.43583 63.92966
  250 72.98227 69.43583 63.92966
```

We tabulated the scores for the Sensitivity-PAUC accroding to the identified factors.

```
> with(pmf.intensity, tapply(TPPAUC, list(Length = Length, Theta = Theta,
+     Scale = Scale, Measure = Measure), mean))

, , Scale = T, Measure = canberra

      Theta
Length     0.5        1        2
  0   87.4013 96.42514 97.06761
  250 87.4013 96.42514 97.06761

, , Scale = N, Measure = canberra

      Theta
Length      0.5        1        2
  0   93.86073 96.97867 97.18447
  250 93.86073 96.97867 97.18447

, , Scale = S, Measure = canberra

      Theta
Length      0.5        1        2
  0   93.86073 96.97867 97.18447
  250 93.86073 96.97867 97.18447

, , Scale = Z, Measure = canberra

      Theta
Length      0.5        1        2
  0   38.48115 45.50660 49.25484
  250 16.11684 16.97443 17.43198
```

```
, , Scale = T, Measure = simindex

      Theta
Length      0.5        1        2
   0    95.09065 96.77991 97.09797
   250 95.09065 96.77991 97.09797

, , Scale = N, Measure = simindex

      Theta
Length      0.5        1        2
   0    96.48608 97.11496 97.21018
   250 96.48608 97.11496 97.21018

, , Scale = S, Measure = simindex

      Theta
Length      0.5        1        2
   0    96.48608 97.11496 97.21018
   250 96.48608 97.11496 97.21018

, , Scale = Z, Measure = simindex

      Theta
Length      0.5        1        2
   0    34.84144 37.28120 39.03578
   250 16.23256 16.36793 16.43746

, , Scale = T, Measure = manhattan

      Theta
Length     0.5        1       2
   0    89.6942 98.07231 98.1431
   250 89.6942 98.07231 98.1431

, , Scale = N, Measure = manhattan

      Theta
Length      0.5        1        2
   0    54.68255 66.73135 72.44122
   250 54.68255 66.73135 72.44122

, , Scale = S, Measure = manhattan

      Theta
Length      0.5        1        2
   0    36.82509 44.65892 49.46728
   250 54.68255 66.73135 72.44122

, , Scale = Z, Measure = manhattan
```

```
      Theta
Length      0.5        1        2
   0   35.65736 43.97521 48.98560
   250 53.06032 64.99398 70.63585

, , Scale = T, Measure = euclidean

      Theta
Length      0.5        1        2
   0   41.28334 62.52455 72.43233
   250 41.28334 62.52455 72.43233

, , Scale = N, Measure = euclidean

      Theta
Length      0.5        1        2
   0   90.84943 97.66087 97.52182
   250 90.84943 97.66087 97.52182

, , Scale = S, Measure = euclidean

      Theta
Length      0.5        1        2
   0   47.99020 59.70180 64.16669
   250 90.84943 97.66087 97.52182

, , Scale = Z, Measure = euclidean

      Theta
Length      0.5        1        2
   0   45.50869 58.28360 62.76938
   250 89.47009 97.44522 97.33485

, , Scale = T, Measure = dotprod

      Theta
Length      0.5        1        2
   0   95.76839 95.76839 95.76839
   250 95.76839 95.76839 95.76839

, , Scale = N, Measure = dotprod

      Theta
Length      0.5        1        2
   0   97.33349 97.33349 97.33349
   250 97.33349 97.33349 97.33349

, , Scale = S, Measure = dotprod
```

```
      Theta
Length      0.5        1        2
   0   95.22259 95.22259 95.22259
   250 97.33349 97.33349 97.33349

, , Scale = Z, Measure = dotprod

      Theta
Length      0.5        1        2
   0   95.68476 93.80895 90.00882
   250 97.25269 97.25252 96.96779

, , Scale = T, Measure = cov

      Theta
Length      0.5        1        2
   0   93.82328 89.78988 84.83983
   250 93.82328 89.78988 84.83983

, , Scale = N, Measure = cov

      Theta
Length      0.5        1       2
   0   95.13124 92.34505 88.6023
   250 95.13124 92.34505 88.6023

, , Scale = S, Measure = cov

      Theta
Length      0.5        1        2
   0   96.80908 96.60766 96.01751
   250 96.80908 96.60766 96.01751

, , Scale = Z, Measure = cov

      Theta
Length      0.5        1        2
   0   96.79598 96.19836 95.13018
   250 96.79598 96.19836 95.13018

, , Scale = T, Measure = soai

      Theta
Length     0.5        1        2
   0   92.6936 93.98806 94.68696
   250 92.6936 93.98806 94.68696

, , Scale = N, Measure = soai

      Theta
```

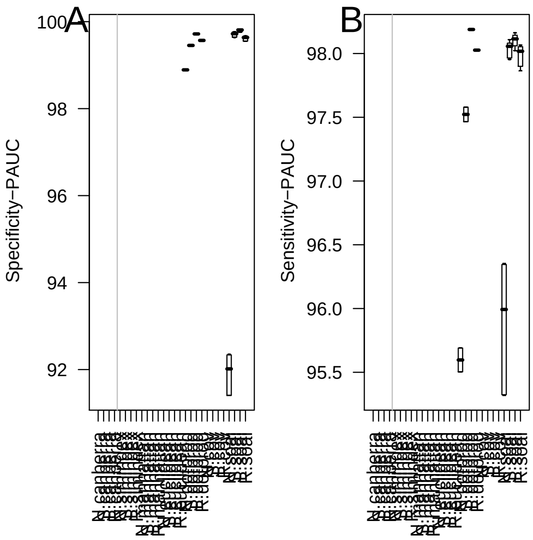
```
Length      0.5        1        2
   0   97.30101 97.54727 97.66245
   250 97.30101 97.54727 97.66245

, , Scale = S, Measure = soai

      Theta
Length      0.5        1        2
   0   97.30101 97.54727 97.66245
   250 97.30101 97.54727 97.66245

, , Scale = Z, Measure = soai

      Theta
Length      0.5        1        2
   0   85.93958 81.75777 70.59478
   250 85.93958 81.75777 70.59478
```

By analysing the table we identify the dotproduct measure computed on vector norm scaled data as having small variance and relatively high scores. Also the euclidean and manhatten distances perform well but only with a sensible parameter choice.

```
> intsoai <- pmf.intensity[(pmf.intensity$Measure == "soai" | pmf.intensity$Measure ==
+     "dotprod") & pmf.intensity$Scale == "N", ]
> intdp <- pmf.intensity[pmf.intensity$Measure == "soai" & pmf.intensity$Scale ==
+     "S", ]
> lmdp <- lm(FPPAUC ~ Length + Theta + Trans + Noncross + Weight,
+     data = intdp)
> anova(lmdp)

Analysis of Variance Table

Response: FPPAUC
          Df    Sum Sq   Mean Sq   F value    Pr(>F)
Length     1 2.935e-27 2.935e-27 1.177e-25    1.0000
Theta      2      1.62      0.81    32.526 2.834e-11 ***
Trans      3   1091.81    363.94 14599.058 < 2.2e-16 ***
Noncross   1 9.046e-10 9.046e-10 3.629e-08    0.9998
Weight     1 1.034e-07 1.034e-07 4.148e-06    0.9984
Residuals 87      2.17      0.02
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(lmdp)$adj.r

[1] 0.9978384
```

By the anova anlysis above we identify the intensity transformation as the only parameter influencing the DP measure. We can how the factor transformation influences the performance of the measures sum of agrein intensities and dot product on the boxplot below.

```
> par(mfrow = c(1, 2))
> par(mar = c(6, 4, 1, 1))
> boxplot(FPPAUC ~ Trans + Measure, data = intsoai, main = "",
+     las = 2)
> mtext("Specificity-PAUC", side = 2, line = 3)
> mtext("A", side = 3, line = -1, at = -3, cex = 2)
> abline(v = 4.5, col = "gray")
> boxplot(TPPAUC ~ Trans + Measure, data = intsoai, main = "",
+     las = 2)
> mtext("Sensitivity-PAUC", side = 2, line = 3)
> mtext("B", side = 3, line = -1, at = -3, cex = 2)
> abline(v = 4.5, col = "gray")
```



The boxplot shows that the log transformation gives highest sensitivities given small FP-rates as well highest specificities given high sensitivities.

1 Analysing the MS/MS dataset

We are going to analyse the MS/MS dataset in a similar way like the PMF dataset and will discuss the differences. The MSMS dataset we use to examine if the conclusion drawn anylysing the PMF dataset can be generalized.

1.1 Binary measure

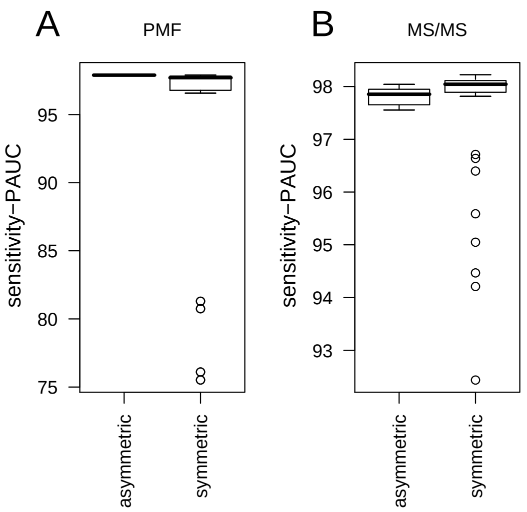
```
> data("msms.binary")
> range(msms.binary$FPPAUC)

[1] 87.67470 99.87236

> range(msms.binary$TPPAUC)

[1] 92.43841 98.22282
```

```
> data("msms.intensity")
> range(msms.intensity$FPPAUC)
```

```
[1]  3.742473 99.981285
```

We start by comparing the performance of the measures having a small variance and performing well in case of the PMF dataset with the performance of all other measures. The left panel compares the FM mallows statistics with the other binary measures. The right panel compares the performance of the dot-product measure with the performance of the other intensity based measures.

```
> par(mfrow = c(1, 2))
> idfm <- rep("symmetric", length(pmf.binary$Measure))
> idfm[which(pmf.binary$Measure == "fm" | pmf.binary$Measure ==
+     "gower")] <- "asymmetric"
> par(mar = c(7, 4, 3, 1))
> boxplot(TPPAUC ~ as.factor(idfm), data = pmf.binary, las = 2,
+     main = "")
> mtext("PMF", side = 3, line = 1)
> mtext("sensitivity-PAUC", side = 2, line = 2.5, cex = 1.2)
> mtext("A", side = 3, line = 1, at = 0, cex = 2)
> idfm <- rep("symmetric", length(msms.binary$Measure))
> idfm[which(msms.binary$Measure == "fm" | msms.binary$Measure ==
+     "gower")] <- "asymmetric"
> par(mar = c(7, 4, 3, 1))
> boxplot(TPPAUC ~ as.factor(idfm), data = msms.binary, las = 2,
+     main = "")
> mtext("MS/MS", side = 3, line = 1)
> mtext("sensitivity-PAUC", side = 2, line = 2.5, cex = 1.2)
> mtext("B", side = 3, line = 1, at = 0, cex = 2)
```




```
> par(mar = c(7, 4, 3, 1))
> ind <- rep("other CP", length(msms.intensity$Measure))
```

```
> ind[msms.intensity$Measure == "dotprod" & msms.intensity$Scale ==
+     "N" & msms.intensity$Trans == "L"] <- "dot-product"
> ind[msms.intensity$Measure == "dotprod"] <- "dot-product"
> boxplot(TPPAUC ~ as.factor(ind), data = msms.intensity, las = 2,
+     main = "", ylim = c(90, 100))
> mtext("intensity", side = 3, line = 1)
> mtext("B", side = 3, line = 1, at = 0, cex = 2)
> mtext("sensitivity-PAUC", side = 2, line = 2.5, cex = 1.2)
```

In case of the MS/MS data the fowlkes mallows statistics does not perform unambiguously better than the other binary measures.

Anova analysis of MSMS dataset

1.2 Binary Measures

Next we apply the same anova model to the MS/MS data as we have applied it to tehe PMF data.

Specificity PAUC – FACTORS

```
> tplm <- lm(FPPAUC ~ Measure + Theta + Length + Weight + Noncross +
+     Length, data = msms.binary)
> summary(tplm)$adj.r

[1] 0.1950604

> anova(tplm)

Analysis of Variance Table

Response: FPPAUC
          Df  Sum Sq Mean Sq F value    Pr(>F)
Measure    3  13.221   4.407  1.7135 0.1701342
Theta      2  43.482  21.741  8.4531 0.0004417 ***
Length     1  19.320  19.320  7.5117 0.0074384 **
Weight     1   0.090   0.090  0.0350 0.8520076
Noncross   1   3.672   3.672  1.4278 0.2353689
Residuals 87 223.761   2.572
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> apply(anova(tplm), 2, sum)

      Df   Sum Sq  Mean Sq  F value   Pr(>F)
 95.0000 303.5473  51.8024       NA       NA

> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

            Df Sum Sq Mean Sq F value Pr(>F)
Measure    3.2    4.4     8.5      NA     NA
```

```
Theta       2.1   14.3    42.0       NA      NA
Length      1.1    6.4    37.3       NA      NA
Weight      1.1    0.0     0.2       NA      NA
Noncross    1.1    1.2     7.1       NA      NA
Residuals 91.6   73.7     5.0       NA      NA
```

Sensitivity PAUC – FACTORS

```
> tplm <- lm(TPPAUC ~ Measure + Theta + Length + Weight + Noncross +
+     Length, data = msms.binary)
> summary(tplm)$adj.r
```

```
[1] 0.1849813
```

```
> round(anova(tplm))
```

```
          Df Sum Sq Mean Sq F value Pr(>F)
Measure    3      2       1       1 <2e-16 ***
Theta      2     12       6       9 <2e-16 ***
Length     1      4       4       7 <2e-16 ***
Weight     1      0       0       0      1
Noncross   1      0       0       1 <2e-16 ***
Residuals 87     54       1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> round(apply(anova(tplm), 2, sum))
```

```
     Df  Sum Sq Mean Sq F value  Pr(>F)
     95      73      12      NA      NA
```

```
> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Measure     3.2    2.7     5.5       NA      NA
Theta       2.1   15.9    48.2       NA      NA
Length      1.1    5.8    35.4       NA      NA
Weight      1.1    0.4     2.2       NA      NA
Noncross    1.1    0.6     3.5       NA      NA
Residuals 91.6   74.6     5.2       NA      NA
```

Specificity PAUC – FACTORS

```
> tplm <- lm(FPPAUC ~ Measure + Theta + Length + Measure * Theta +
+     Measure * Length + Theta * Length + Measure * Theta * Length,
+     data = msms.binary)
> summary(tplm)$adj.r
```

```
[1] 0.7238965
```

```
> round(anova(tplm))
```

```
                    Df Sum Sq Mean Sq F value    Pr(>F)    
Measure              3     13       4       5 < 2.2e-16 ***
Theta                2     43      22      25 < 2.2e-16 ***
Length               1     19      19      22 < 2.2e-16 ***
Measure:Theta        6     50       8       9 < 2.2e-16 ***
Measure:Length       3     22       7       8 < 2.2e-16 ***
Theta:Length         2     44      22      25 < 2.2e-16 ***
Measure:Theta:Length  6     48       8       9 < 2.2e-16 ***
Residuals           72     64       1                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> apply(anova(tplm), 2, sum)
```

```
      Df   Sum Sq  Mean Sq  F value   Pr(>F) 
 95.0000 303.5473  91.9542       NA       NA 
```

```
> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
                      Df Sum Sq Mean Sq F value Pr(>F)
Measure              3.2    4.4     4.8      NA     NA
Theta                2.1   14.3    23.6      NA     NA
Length               1.1    6.4    21.0      NA     NA
Measure:Theta        6.3   16.3     9.0      NA     NA
Measure:Length       3.2    7.3     8.0      NA     NA
Theta:Length         2.1   14.4    23.8      NA     NA
Measure:Theta:Length  6.3   16.0     8.8      NA     NA
Residuals           75.8   20.9     1.0      NA     NA
```

Sensitivity PAUC – FINAL

```
> tplm <- lm(TPPAUC ~ Measure + Theta + Length + Measure * Theta +
+     Measure * Length + Theta * Length + Measure * Theta * Length,
+     data = msms.binary)
> summary(tplm)$adj.r
```

```
[1] 0.7378059
```

```
> round(anova(tplm))
```

```
                    Df Sum Sq Mean Sq F value    Pr(>F)    
Measure              3      2       1       3 < 2.2e-16 ***
Theta                2     12       6      29 < 2.2e-16 ***
Length               1      4       4      21 < 2.2e-16 ***
Measure:Theta        6     13       2      10 < 2.2e-16 ***
Measure:Length       3      5       2       8 < 2.2e-16 ***
Theta:Length         2     11       6      28 < 2.2e-16 ***
Measure:Theta:Length  6     12       2      10 < 2.2e-16 ***
Residuals           72     14       0                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> apply(anova(tplm), 2, sum)

      Df   Sum Sq  Mean Sq  F value   Pr(>F)
95.00000 72.85633 22.23895       NA       NA

> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

                       Df Sum Sq Mean Sq F value Pr(>F)
Measure               3.2    2.7     3.0      NA     NA
Theta                 2.1   15.9    26.0      NA     NA
Length                1.1    5.8    19.1      NA     NA
Measure:Theta         6.3   17.2     9.4      NA     NA
Measure:Length        3.2    6.5     7.1      NA     NA
Theta:Length          2.1   15.6    25.6      NA     NA
Measure:Theta:Length  6.3   16.4     8.9      NA     NA
Residuals            75.8   19.9     0.9      NA     NA
```

We see that the old model is not performing as well as in case of the PMF data. A model which includes computing the noncrossing matching performs much better.

```
> tplm <- lm(TPPAUC ~ Measure * Theta * Length * Noncross, data = msms.binary)
> summary(tplm)$adj

[1] 0.908328

> anova(tplm)

Analysis of Variance Table

Response: TPPAUC
                              Df  Sum Sq Mean Sq F value    Pr(>F)
Measure                        3  1.9891  0.6630  9.4311 5.224e-05 ***
Theta                          2 11.5674  5.7837 82.2667 3.101e-16 ***
Length                         1  4.2395  4.2395 60.3030 4.988e-10 ***
Noncross                       1  0.4206  0.4206  5.9826 0.0181639 *
Measure:Theta                  6 12.5397  2.0900 29.7273 1.408e-14 ***
Measure:Length                 3  4.7251  1.5750 22.4032 3.254e-09 ***
Theta:Length                   2 11.4009  5.7005 81.0829 4.058e-16 ***
Measure:Noncross               3  1.0371  0.3457  4.9171 0.0046521 **
Theta:Noncross                 2  1.8214  0.9107 12.9536 3.172e-05 ***
Length:Noncross                1  0.9581  0.9581 13.6279 0.0005692 ***
Measure:Theta:Length           6 11.9168  1.9861 28.2507 3.595e-14 ***
Measure:Theta:Noncross         6  1.9925  0.3321  4.7235 0.0007470 ***
Measure:Length:Noncross        3  1.0254  0.3418  4.8618 0.0049388 **
Theta:Length:Noncross          2  1.8333  0.9167 13.0385 3.002e-05 ***
Measure:Theta:Length:Noncross  6  2.0147  0.3358  4.7762 0.0006850 ***
Residuals                     48  3.3746  0.0703
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> dr <- anova(tplm)
> ddd <- apply(anova(tplm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)
```

```
                                Df Sum Sq Mean Sq F value Pr(>F)
Measure                        3.2    2.7     2.5      NA     NA
Theta                          2.1   15.9    21.7      NA     NA
Length                         1.1    5.8    15.9      NA     NA
Noncross                       1.1    0.6     1.6      NA     NA
Measure:Theta                  6.3   17.2     7.8      NA     NA
Measure:Length                 3.2    6.5     5.9      NA     NA
Theta:Length                   2.1   15.6    21.4      NA     NA
Measure:Noncross               3.2    1.4     1.3      NA     NA
Theta:Noncross                 2.1    2.5     3.4      NA     NA
Length:Noncross                1.1    1.3     3.6      NA     NA
Measure:Theta:Length           6.3   16.4     7.4      NA     NA
Measure:Theta:Noncross         6.3    2.7     1.2      NA     NA
Measure:Length:Noncross        3.2    1.4     1.3      NA     NA
Theta:Length:Noncross          2.1    2.5     3.4      NA     NA
Measure:Theta:Length:Noncross  6.3    2.8     1.3      NA     NA
Residuals                     50.5    4.6     0.3      NA     NA
```

Tabulate the data according to identified factors.

```
> with(msms.binary, tapply(TPPAUC, list(Length = Length, Theta = Theta,
+     Noncross = Noncross, Measure = Measure), mean))
```

```
, , Noncross = A, Measure = fm

      Theta
Length      0.5       1        2
   0   97.92753 97.9485 97.95956
   250 97.92753 97.9485 97.95956

, , Noncross = O, Measure = fm

      Theta
Length      0.5        1        2
   0   97.84375 97.86718 97.87783
   250 97.84375 97.86718 97.87783

, , Noncross = A, Measure = gower

      Theta
Length      0.5        1        2
   0   97.75744 97.75744 97.75744
   250 97.75744 97.75744 97.75744

, , Noncross = O, Measure = gower

      Theta
```

```
Length      0.5        1        2
   0   97.70388 97.70388 97.70388
   250 97.70388 97.70388 97.70388

, , Noncross = A, Measure = hg

      Theta
Length      0.5        1        2
   0   98.09396 98.05886 98.03635
   250 98.13553 98.18004 93.32485

, , Noncross = O, Measure = hg

      Theta
Length      0.5        1        2
   0   98.04063 97.98980 97.96519
   250 98.05277 98.12767 96.11346

, , Noncross = A, Measure = rmi

      Theta
Length      0.5        1        2
   0   97.98795 97.98937 97.98952
   250 98.19244 98.12313 94.75886

, , Noncross = O, Measure = rmi

      Theta
Length      0.5        1        2
   0   97.91691 97.92064 97.92132
   250 98.12738 98.13180 96.55549
```

We see that the binary measures computed with noncrossing matching perform better than they associates without. Furthermore Huberts Gamma computed with $M_{00} = 0$ can be recognized as the best performing measure. The same analysis is repeated for the Sp-PAUC.

```
> boxplot(FPPAUC ~ Measure, data = msms.binary)
> tplm <- lm(FPPAUC ~ Measure * Theta * Length * Noncross, data = msms.binary)
> summary(tplm)$adj

[1] 0.9215654

> anova(tplm)

Analysis of Variance Table

Response: FPPAUC
                              Df Sum Sq Mean Sq F value    Pr(>F)
Measure                        3 13.221   4.407 17.5850 7.724e-08 ***
Theta                          2 43.482  21.741 86.7507 < 2.2e-16 ***
```

```
Length                         1 19.320  19.320 77.0890 1.498e-11 ***
Noncross                       1  3.672   3.672 14.6530 0.0003739 ***
Measure:Theta                  6 49.601   8.267 32.9858 2.007e-15 ***
Measure:Length                 3 22.190   7.397 29.5133 5.799e-11 ***
Theta:Length                   2 43.715  21.857 87.2143 < 2.2e-16 ***
Measure:Noncross               3  4.559   1.520  6.0634 0.0013831 **
Theta:Noncross                 2  8.038   4.019 16.0367 4.635e-06 ***
Length:Noncross                1  4.039   4.039 16.1164 0.0002082 ***
Measure:Theta:Length           6 48.499   8.083 32.2532 3.069e-15 ***
Measure:Theta:Noncross         6  9.233   1.539  6.1405 7.861e-05 ***
Measure:Length:Noncross        3  4.632   1.544  6.1613 0.0012501 **
Theta:Length:Noncross          2  8.036   4.018 16.0323 4.647e-06 ***
Measure:Theta:Length:Noncross  6  9.280   1.547  6.1713 7.499e-05 ***
Residuals                     48 12.030   0.251
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabulate the data according to identified factors.

```
> with(msms.binary, tapply(TPPAUC, list(Length = Length, Theta = Theta,
+     Noncross = Noncross, Measure = Measure), mean))
```

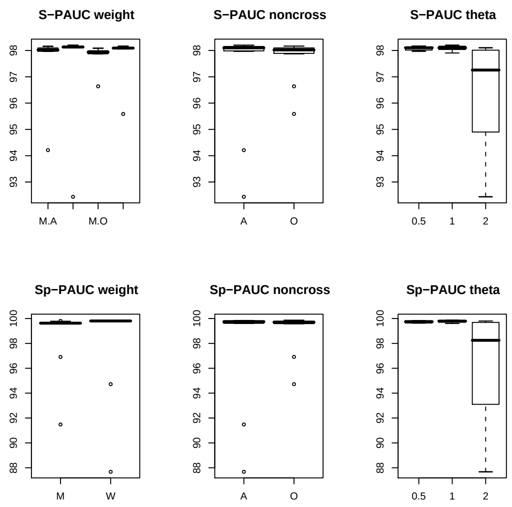
```
, , Noncross = A, Measure = fm

      Theta
Length      0.5       1        2
   0   97.92753 97.9485 97.95956
   250 97.92753 97.9485 97.95956

, , Noncross = O, Measure = fm

      Theta
Length      0.5        1        2
   0   97.84375 97.86718 97.87783
   250 97.84375 97.86718 97.87783

, , Noncross = A, Measure = gower

      Theta
Length      0.5        1        2
   0   97.75744 97.75744 97.75744
   250 97.75744 97.75744 97.75744

, , Noncross = O, Measure = gower

      Theta
Length      0.5        1        2
   0   97.70388 97.70388 97.70388
   250 97.70388 97.70388 97.70388

, , Noncross = A, Measure = hg
```

```
      Theta
Length      0.5        1        2
   0   98.09396 98.05886 98.03635
   250 98.13553 98.18004 93.32485

, , Noncross = O, Measure = hg

      Theta
Length      0.5        1        2
   0   98.04063 97.98980 97.96519
   250 98.05277 98.12767 96.11346

, , Noncross = A, Measure = rmi

      Theta
Length      0.5        1        2
   0   97.98795 97.98937 97.98952
   250 98.19244 98.12313 94.75886

, , Noncross = O, Measure = rmi

      Theta
Length      0.5        1        2
   0   97.91691 97.92064 97.92132
   250 98.12738 98.13180 96.55549
```

Again hg performs best.

Further we analysed how the HG measure depends on factors like e.g. Weighting of mass measurement error, noncrossing matching and theta.

```
> msms.binaryhg <- msms.binary[msms.binary$Measure == "hg", ]
> par(mfcol = c(2, 3))
> boxplot(TPPAUC ~ Weight + Noncross, data = msms.binaryhg, main = "S-PAUC weight")
> boxplot(FPPAUC ~ Weight, data = msms.binaryhg, main = "Sp-PAUC weight")
> boxplot(TPPAUC ~ Noncross, data = msms.binaryhg, main = "S-PAUC noncross")
> boxplot(FPPAUC ~ Noncross, data = msms.binaryhg, main = "Sp-PAUC noncross")
> boxplot(TPPAUC ~ Theta, data = msms.binaryhg, main = "S-PAUC theta")
> boxplot(FPPAUC ~ Theta, data = msms.binaryhg, main = "Sp-PAUC theta")
```



The boxplots reveal that weighting of match accuracy and resolving unambigous matches by computing the noncrossing matching increases the performance of the HG measure. Furthermore the optimal choice of the theta is 2.

1.2.1 MS/MS-Intensity based measures

Specificity PAUC – FACTORS

```
> intlm <- lm(FPPAUC ~ Measure + Trans + Scale + Theta + Weight +
+     Noncross + Length, data = msms.intensity)
> anova(intlm)
```

```
Analysis of Variance Table

Response: FPPAUC
            Df  Sum Sq Mean Sq  F value    Pr(>F)
Measure      6  283609   47268 113.4517 < 2.2e-16 ***
Trans        3   11843    3948   9.4749 3.173e-06 ***
Scale        3  454533  151511 363.6517 < 2.2e-16 ***
Theta        2   25668   12834  30.8034 5.946e-14 ***
Weight       1    4606    4606  11.0563 0.0008958 ***
Noncross     1     733     733   1.7601 0.1847308
Length       1    5343    5343  12.8235 0.0003484 ***
Residuals 2670 1112422     417
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(intlm)$adj.r
```

```
[1] 0.4104012
```

```
> apply(anova(intlm), 2, sum)
```

```
       Df     Sum Sq    Mean Sq    F value     Pr(>F)
   2687.0 1898757.6  226659.8         NA         NA
```

```
> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

            Df Sum Sq Mean Sq F value Pr(>F)
Measure    0.2   14.9    20.9      NA     NA
Trans      0.1    0.6     1.7      NA     NA
Scale      0.1   23.9    66.8      NA     NA
Theta      0.1    1.4     5.7      NA     NA
Weight     0.0    0.2     2.0      NA     NA
Noncross   0.0    0.0     0.3      NA     NA
Length     0.0    0.3     2.4      NA     NA
Residuals 99.4   58.6     0.2      NA     NA
```

Sensitivity PAUC – FACTORS

```
> intlm <- lm(TPPAUC ~ Measure + Trans + Scale + Theta + Weight +
+     Noncross + Length, data = msms.intensity)
> anova(intlm)

Analysis of Variance Table

Response: TPPAUC
            Df Sum Sq Mean Sq  F value    Pr(>F)
Measure      6 152255   25376 114.2873 < 2.2e-16 ***
Trans        3   2869     956   4.3072  0.004874 **
Scale        3 254688   84896 382.3532 < 2.2e-16 ***
Theta        2   9241    4620  20.8094 1.077e-09 ***
Weight       1   1971    1971   8.8789  0.002911 **
Noncross     1     74      74   0.3319  0.564578
Length       1   1340    1340   6.0334  0.014101 *
Residuals 2670 592834     222
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(intlm)$adj.r

[1] 0.4123652

> apply(anova(intlm), 2, sum)

       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 1015271.6  119455.3        NA        NA

> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

            Df Sum Sq Mean Sq F value Pr(>F)
Measure    0.2   15.0    21.2      NA     NA
Trans      0.1    0.3     0.8      NA     NA
Scale      0.1   25.1    71.1      NA     NA
```

```
Theta       0.1    0.9     3.9      NA     NA
Weight      0.0    0.2     1.7      NA     NA
Noncross    0.0    0.0     0.1      NA     NA
Length      0.0    0.1     1.1      NA     NA
Residuals 99.4   58.4     0.2      NA     NA
```

Specificity PAUC – FINAL

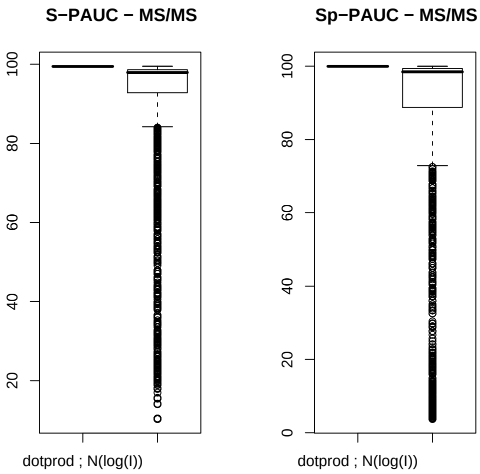
```
> intlm <- lm(FPPAUC ~ Measure + Scale + Theta + Length + Measure:Scale +
+     Measure:Theta + Measure:Length, data = msms.intensity)
> anova(intlm)

Analysis of Variance Table

Response: FPPAUC
                 Df Sum Sq Mean Sq  F value    Pr(>F)
Measure           6 283609   47268  574.207 < 2.2e-16 ***
Scale             3 454533  151511 1840.530 < 2.2e-16 ***
Theta             2  25668   12834  155.904 < 2.2e-16 ***
Length            1   5343    5343   64.903 1.179e-15 ***
Measure:Scale    18 824919   45829  556.721 < 2.2e-16 ***
Measure:Theta    12  70199    5850   71.064 < 2.2e-16 ***
Measure:Length    6  17247    2874   34.918 < 2.2e-16 ***
Residuals      2639 217240      82
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(intlm)$adj.r

[1] 0.8835072

> apply(anova(intlm), 2, sum)

       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 1898757.6  271591.3        NA        NA

> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

                 Df Sum Sq Mean Sq F value Pr(>F)
Measure         0.2   14.9    17.4      NA     NA
Scale           0.1   23.9    55.8      NA     NA
Theta           0.1    1.4     4.7      NA     NA
Length          0.0    0.3     2.0      NA     NA
Measure:Scale   0.7   43.4    16.9      NA     NA
Measure:Theta   0.4    3.7     2.2      NA     NA
Measure:Length  0.2    0.9     1.1      NA     NA
Residuals      98.2   11.4     0.0      NA     NA
```

Sensitivity PAUC – FINAL

```
> intlm <- lm(TPPAUC ~ Measure + Scale + Theta + Length + Measure:Scale +
+     Measure:Theta + Measure:Length, data = msms.intensity)
> anova(intlm)

Analysis of Variance Table

Response: TPPAUC
                 Df Sum Sq Mean Sq  F value    Pr(>F)
Measure           6 152255   25376  618.221 < 2.2e-16 ***
Scale             3 254688   84896 2068.286 < 2.2e-16 ***
Theta             2   9241    4620  112.565 < 2.2e-16 ***
Length            1   1340    1340   32.637 1.235e-08 ***
Measure:Scale    18 446982   24832  604.981 < 2.2e-16 ***
Measure:Theta    12  23460    1955   47.629 < 2.2e-16 ***
Measure:Length    6  18984    3164   77.085 < 2.2e-16 ***
Residuals      2639 108322      41
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(intlm)$adj.r

[1] 0.891367

> apply(anova(intlm), 2, sum)

       Df    Sum Sq   Mean Sq   F value    Pr(>F)
   2687.0 1015271.6  146224.2        NA        NA

> dr <- anova(intlm)
> ddd <- apply(anova(intlm), 2, sum)
> round(sweep(as.matrix(dr), 2, ddd, "/") * 100, digits = 1)

                 Df Sum Sq Mean Sq F value Pr(>F)
Measure         0.2   15.0    17.4      NA     NA
Scale           0.1   25.1    58.1      NA     NA
Theta           0.1    0.9     3.2      NA     NA
Length          0.0    0.1     0.9      NA     NA
Measure:Scale   0.7   44.0    17.0      NA     NA
Measure:Theta   0.4    2.3     1.3      NA     NA
Measure:Length  0.2    1.9     2.2      NA     NA
Residuals      98.2   10.7     0.0      NA     NA
```

Again we first prove if the result obtained for the PMF data can be generalized to the MS/MS data.

```
> ind <- rep("other measure", length(msms.intensity$Measure))
> ind[msms.intensity$Measure == "dotprod" & msms.intensity$Scale ==
+     "N" & msms.intensity$Trans == "L"] <- "dotprod ; N(log(I))"
> par(mfrow = c(1, 2))
> boxplot(TPPAUC ~ as.factor(ind), data = msms.intensity, main = "S-PAUC - MS/MS")
> boxplot(FPPAUC ~ as.factor(ind), data = msms.intensity, main = "Sp-PAUC - MS/MS")
```



This time the observation done using the PMF data can be generalized to the MS/MS data. Because we were interested to identify the other measures which can be used to classify the data we tabulated the scores according to theta, scaling, length and the measures.

```
> intlm <- lm(FPPAUC ~ Measure + Scale + Theta + Length + Measure:Scale +
+     Measure:Theta + Measure:Length, data = msms.intensity)
> anova(intlm)

Analysis of Variance Table

Response: FPPAUC
                 Df Sum Sq Mean Sq  F value    Pr(>F)    
Measure           6 283609   47268  574.207 < 2.2e-16 ***
Scale             3 454533  151511 1840.530 < 2.2e-16 ***
Theta             2  25668   12834  155.904 < 2.2e-16 ***
Length            1   5343    5343   64.903 1.179e-15 ***
Measure:Scale    18 824919   45829  556.721 < 2.2e-16 ***
Measure:Theta    12  70199    5850   71.064 < 2.2e-16 ***
Measure:Length    6  17247    2874   34.918 < 2.2e-16 ***
Residuals      2639 217240      82                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(intlm)$adj.r

[1] 0.8835072

> with(msms.intensity, tapply(FPPAUC, list(Length = Length, Theta = Theta,
+     Scale = Scale, Measure = Measure), mean))

, , Scale = T, Measure = canberra

      Theta
```

```
Length      0.5        1        2
   0   90.50912 97.76038 98.38236
   250 90.50912 97.76038 98.38236

, , Scale = N, Measure = canberra

      Theta
Length      0.5        1       2
   0   96.93366 98.77676 98.7238
   250 96.93366 98.77676 98.7238

, , Scale = S, Measure = canberra

      Theta
Length      0.5        1       2
   0   96.93366 98.77676 98.7238
   250 96.93366 98.77676 98.7238

, , Scale = Z, Measure = canberra

      Theta
Length       0.5         1         2
   0   10.895818 13.267293 15.003216
   250  6.557513  6.781328  6.903368

, , Scale = T, Measure = simindex

      Theta
Length      0.5        1      2
   0   97.23546 98.34243 98.523
   250 97.23546 98.34243 98.523

, , Scale = N, Measure = simindex

      Theta
Length     0.5        1        2
   0   98.6569 98.78721 98.70548
   250 98.6569 98.78721 98.70548

, , Scale = S, Measure = simindex

      Theta
Length     0.5        1        2
   0   98.6569 98.78721 98.70548
   250 98.6569 98.78721 98.70548

, , Scale = Z, Measure = simindex

      Theta
Length      0.5        1         2
```

```
   0   9.293497 9.909243 10.416333
   250 5.270600 5.281156  5.288262

, , Scale = T, Measure = manhattan

      Theta
Length      0.5        1       2
   0   84.90806 99.63053 99.9417
   250 84.90806 99.63053 99.9417

, , Scale = N, Measure = manhattan

      Theta
Length      0.5        1        2
   0   64.01813 82.40765 89.32846
   250 64.01813 82.40765 89.32846

, , Scale = S, Measure = manhattan

      Theta
Length      0.5        1        2
   0   42.28481 59.21963 68.55283
   250 64.01813 82.40765 89.32846

, , Scale = Z, Measure = manhattan

      Theta
Length      0.5        1        2
   0   28.95919 44.71444 54.62749
   250 56.29775 76.75337 85.29903

, , Scale = T, Measure = euclidean

      Theta
Length      0.5        1        2
   0   34.31829 73.02366 87.94384
   250 34.31829 73.02366 87.94384

, , Scale = N, Measure = euclidean

      Theta
Length      0.5        1        2
   0   86.13116 99.88037 99.83952
   250 86.13116 99.88037 99.83952

, , Scale = S, Measure = euclidean

      Theta
Length      0.5        1        2
   0   69.66885 92.44349 95.00940
```

```
   250 86.13116 99.88037 99.83952

, , Scale = Z, Measure = euclidean

      Theta
Length      0.5        1        2
   0    52.85492 81.78122 86.69784
   250 85.87138 99.87904 99.83569

, , Scale = T, Measure = dotprod

      Theta
Length     0.5       1       2
   0    98.9814 98.9814 98.9814
   250 98.9814 98.9814 98.9814

, , Scale = N, Measure = dotprod

      Theta
Length      0.5        1        2
   0    99.88292 99.88292 99.88292
   250 99.88292 99.88292 99.88292

, , Scale = S, Measure = dotprod

      Theta
Length      0.5        1        2
   0    95.17252 95.17252 95.17252
   250 99.88292 99.88292 99.88292

, , Scale = Z, Measure = dotprod

      Theta
Length      0.5        1        2
   0    99.54657 98.39198 95.06786
   250 99.89702 99.90375 99.90253

, , Scale = T, Measure = cov

      Theta
Length     0.5      1        2
   0    97.5566 94.342 89.89063
   250 97.5566 94.342 89.89063

, , Scale = N, Measure = cov

      Theta
Length     0.5        1        2
   0    98.6932 96.95327 94.23659
   250 98.6932 96.95327 94.23659
```

```
, , Scale = S, Measure = cov

      Theta
Length      0.5        1        2
   0    99.45862 99.28724 98.85178
   250 99.45862 99.28724 98.85178

, , Scale = Z, Measure = cov

      Theta
Length      0.5        1        2
   0    99.27588 98.53667 97.24777
   250 99.27588 98.53667 97.24777

, , Scale = T, Measure = soai

      Theta
Length      0.5        1        2
   0    96.43898 97.95112 98.64927
   250 96.43898 97.95112 98.64927

, , Scale = N, Measure = soai

      Theta
Length     0.5       1        2
   0    99.5148 99.6997 99.78143
   250 99.5148 99.6997 99.78143

, , Scale = S, Measure = soai

      Theta
Length     0.5       1        2
   0    99.5148 99.6997 99.78143
   250 99.5148 99.6997 99.78143

, , Scale = Z, Measure = soai

      Theta
Length      0.5        1        2
   0    92.05551 88.75732 82.87961
   250 92.05551 88.75732 82.87961

> intlm <- lm(TPPAUC ~ Measure + Scale + Theta + Length + Measure *
+     Scale + Measure * Theta + Measure * Length, data = msms.intensity)
> anova(intlm)

Analysis of Variance Table

Response: TPPAUC
```

```
                 Df Sum Sq Mean Sq  F value    Pr(>F)    
Measure           6 152255   25376  618.221 < 2.2e-16 ***
Scale             3 254688   84896 2068.286 < 2.2e-16 ***
Theta             2   9241    4620  112.565 < 2.2e-16 ***
Length            1   1340    1340   32.637 1.235e-08 ***
Measure:Scale    18 446982   24832  604.981 < 2.2e-16 ***
Measure:Theta    12  23460    1955   47.629 < 2.2e-16 ***
Measure:Length    6  18984    3164   77.085 < 2.2e-16 ***
Residuals      2639 108322      41                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(intlm)$adj.r
```

```
[1] 0.891367
```

```
> with(msms.intensity, tapply(TPPAUC, list(Length = Length, Theta = Theta,
+     Scale = Scale, Measure = Measure), mean))
```

```
, , Scale = T, Measure = canberra

      Theta
Length      0.5        1        2
   0   95.76344 97.64019 98.00106
   250 95.76344 97.64019 98.00106

, , Scale = N, Measure = canberra

      Theta
Length      0.5        1       2
   0   97.59895 98.24352 98.2426
   250 97.59895 98.24352 98.2426

, , Scale = S, Measure = canberra

      Theta
Length      0.5        1       2
   0   97.59895 98.24352 98.2426
   250 97.59895 98.24352 98.2426

, , Scale = Z, Measure = canberra

      Theta
Length      0.5        1        2
   0   41.54274 50.17743 55.22927
   250 25.11656 26.57378 27.34388

, , Scale = T, Measure = simindex

      Theta
Length      0.5       1       2
```

```
   0   97.44322 97.9621 98.10808
   250 97.44322 97.9621 98.10808

, , Scale = N, Measure = simindex

      Theta
Length      0.5        1        2
   0   98.18439 98.28261 98.24748
   250 98.18439 98.28261 98.24748

, , Scale = S, Measure = simindex

      Theta
Length      0.5        1        2
   0   98.18439 98.28261 98.24748
   250 98.18439 98.28261 98.24748

, , Scale = Z, Measure = simindex

      Theta
Length      0.5        1        2
   0   29.83569 32.82933 35.31495
   250 18.64167 18.78237 18.85644

, , Scale = T, Measure = manhattan

      Theta
Length     0.5        1       2
   0   94.4853 98.65037 98.7925
   250 94.4853 98.65037 98.7925

, , Scale = N, Measure = manhattan

      Theta
Length      0.5        1        2
   0   73.61772 87.05038 91.65706
   250 73.61772 87.05038 91.65706

, , Scale = S, Measure = manhattan

      Theta
Length      0.5        1        2
   0   52.10334 66.04717 73.87451
   250 73.61772 87.05038 91.65706

, , Scale = Z, Measure = manhattan

      Theta
Length      0.5        1        2
   0   45.43646 58.34826 66.50964
```

```
   250 69.82565 84.43180 89.62166

, , Scale = T, Measure = euclidean

      Theta
Length      0.5      1        2
   0   64.26425 85.714 92.55688
   250 64.26425 85.714 92.55688

, , Scale = N, Measure = euclidean

      Theta
Length      0.5        1        2
   0   94.65832 98.76085 98.76522
   250 94.65832 98.76085 98.76522

, , Scale = S, Measure = euclidean

      Theta
Length      0.5        1        2
   0   77.58234 92.43829 94.50953
   250 94.65832 98.76085 98.76522

, , Scale = Z, Measure = euclidean

      Theta
Length     0.5        1        2
   0   66.8460 86.27333 89.62940
   250 94.3119 98.80819 98.80822

, , Scale = T, Measure = dotprod

      Theta
Length     0.5       1       2
   0   97.4194 97.4194 97.4194
   250 97.4194 97.4194 97.4194

, , Scale = N, Measure = dotprod

      Theta
Length      0.5        1        2
   0   99.22425 99.22425 99.22425
   250 99.22425 99.22425 99.22425

, , Scale = S, Measure = dotprod

      Theta
Length      0.5        1        2
   0   94.63101 94.63101 94.63101
   250 99.22425 99.22425 99.22425
```

```
, , Scale = Z, Measure = dotprod

      Theta
Length      0.5        1        2
   0    98.92080 98.32497 96.70782
   250 99.20982 99.22335 99.23680

, , Scale = T, Measure = cov

      Theta
Length     0.5        1       2
   0    97.3018 96.95074 95.9306
   250 97.3018 96.95074 95.9306

, , Scale = N, Measure = cov

      Theta
Length      0.5        1        2
   0    98.21523 97.94886 97.29641
   250 98.21523 97.94886 97.29641

, , Scale = S, Measure = cov

      Theta
Length      0.5        1       2
   0    98.57152 98.54664 98.3778
   250 98.57152 98.54664 98.3778

, , Scale = Z, Measure = cov

      Theta
Length      0.5        1      2
   0    98.56075 98.32342 97.922
   250 98.56075 98.32342 97.922

, , Scale = T, Measure = soai

      Theta
Length      0.5        1        2
   0    97.28639 97.95305 98.28481
   250 97.28639 97.95305 98.28481

, , Scale = N, Measure = soai

      Theta
Length     0.5        1        2
   0    98.6569 98.81979 98.89044
   250 98.6569 98.81979 98.89044
```

```
, , Scale = S, Measure = soai

      Theta
Length     0.5        1        2
   0    98.6569 98.81979 98.89044
   250 98.6569 98.81979 98.89044

, , Scale = Z, Measure = soai

      Theta
Length      0.5        1        2
   0    95.52872 92.78444 82.18235
   250 95.52872 92.78444 82.18235
```

The spectral angle measure is the highest scoring one. None of the other measures is able to obtain similar scores. This may be due to the fact that the database search of the MS/MS data is performed using the normalized crosscorelation which has a very similar mathematical property than the spectral angle.

Finally we analyse how factors like intensity transformation, weighting of mass measurement accuracy and computing the noncrossing matching influences the performance of the spectral angle.

```
> intdp <- msms.intensity[(msms.intensity$Measure == "euclidean") &
+     msms.intensity$Scale == "S" & msms.intensity$Trans == "L",
+     ]
> boxplot(TPPAUC ~ Weight * Noncross * Measure * Trans, data = intdp,
+     main = "S-PAUC weight", las = 2)
> par(mar = c(8, 5, 2, 2))
> boxplot(FPPAUC ~ Noncross, data = intdp, main = "S-PAUC weight",
+     las = 2)
> boxplot(TPPAUC ~ Weight, data = intdp, main = "S-PAUC weight",
+     las = 2)

> boxplot(TPPAUC ~ Weight * Noncross * Theta * Measure, data = intdp,
+     main = "S-PAUC weight", las = 2)

> lmdp <- lm(FPPAUC ~ Weight + Noncross + Weight:Noncross, data = intdp)
> anova(lmdp)

Analysis of Variance Table

Response: FPPAUC
                Df  Sum Sq Mean Sq F value Pr(>F)
Weight           1   27.23   27.23  0.4959 0.4894
Noncross         1   52.16   52.16  0.9498 0.3414
Weight:Noncross  1    1.65    1.65  0.0300 0.8643
Residuals       20 1098.25   54.91

> summary(lmdp)$adj.r

[1] -0.07097703
```

```
> par(mfcol = c(2, 3))
> boxplot(TPPAUC ~ Trans, data = intdp, main = "S-PAUC trans")
> boxplot(FPPAUC ~ Trans, data = intdp, main = "Sp-PAUC trans")
> boxplot(FPPAUC ~ Weight, data = intdp, main = "Sp-PAUC weight")
> boxplot(FPPAUC ~ Noncross, data = intdp, main = "S-PAUC noncross")
> boxplot(TPPAUC ~ Noncross, data = intdp, main = "Sp-PAUC noncross")
```