Smacof at 50: A Manual
Part x: smacofRO - Non-metric Smacof for Rank Order Data

Jan de Leeuw - University of California Los Angeles

Started May 12 2024, Version of May 12, 2024

Contents

Note: This is a working manuscript which will be expanded/updated frequently. All suggestions for improvement are welcome. All Rmd, tex, html, pdf, R, and C files are in the public domain. Attribution will be appreciated, but is not required. The files can be found at https://github.com/ deleeuw in the repositories smacofCode, smacofManual, and smacofExamples.

1 Introduction

The smacofRO program is a convential non-metric MDS program designed to handle a partial order over pairs from a single set of objects. Typically the partial order is derived from a symmetric matrix of numerical dissimilarities or from a ranking (with or without ties) of dissimilarities.

2 Data

3 Program

3.1 Parameters

```
smacofRO <- function(data,
                     ndim,
                     xold = NULL,
                     labels = NULL,
                     width = 15,
                     precision = 10,
                     itmax = 1000,
                     eps = 1e-10,
                     verbose = TRUE,
                     kitmax = 5,
                     keps = 1e-10,
                     kverbose = 0,
                     init = 1,
                     ties = 1)
```

3.2 Input

The data are a five column table. Here are the first five rows for the Gruijter example, analyzed below.

```
     i j delta      weight ties
[1,] 7 6  3.20 0.02777778    1
[2,] 2 1  4.08 0.02777778    2
[3,] 3 2  4.59 0.02777778    3
[4,] 6 5  4.60 0.02777778    4
[5,] 8 4  4.67 0.02777778    5
```

The first two columns give the indices of the dissimilarities. Always $1 \leq j < i \leq n$. Column three gives the value of delta, which can be an observed numerical value or a rank number. The data are in non-increasing order within their column. Column four gives the weights, adding up to one over all observations. And the last column codes tieblocks. It is allowed to have are fewer than $\frac{1}{2}n(n-1)$ rows. The missing rows are computationally equivalent to non-missing rows with weight equal to zero.

Typically we create the data by using the utility function, which is part of the smacofRO package. Here delta is a symmetric matrix or an object of class dist. If weights is NULL all weights are assumed to be equal.

```
smacofMakeRankOrderData <-function(delta, weights = NULL, tieblocks = TRUE)
```

3.3 Algorithm

The ALS algorithm alternates a number of Guttman iterations with a single monotone regression (with either one of the three options for ties, cf De Leeuw (1977)). Guttman iterations are normalized explicitly, using $\sum \sum w_{ij} d_{ij}^2(X) = 1$. Monotone regression is done with the pava algorithm from De Leeuw (2017).

3.4 Output
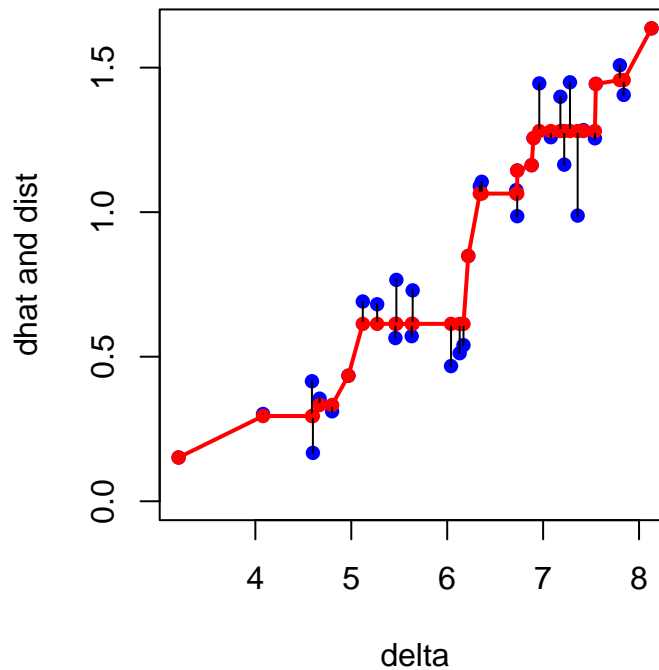
4 Examples

4.1 De Gruijter (1967)

In this example the number of tieblocks is 35, which means there is a single tieblock of two observations, the other 34 observations are untied. We expect that the three different ways of handling ties will not make much of a difference.

```
h1 <- smacofRO(gruijterData, 2, ties = 1, labels = labels, verbose = FALSE)
h2 <- smacofRO(gruijterData, 2, ties = 2, labels = labels, verbose = FALSE)
h3 <- smacofRO(gruijterData, 2, ties = 3, labels = labels, verbose = FALSE)
```

- Ties = 1 uses 88 iterations and stops at stress 0.0042180140
- Ties = 2 uses 65 iterations and stops at stress 0.0042573281
- Ties = 3 uses 69 iterations and stops at stress 0.0040850898

And indeed the solutions are practically the same, although the number of iterations needed for convergence differs somewhat between options. We only show one Shepardplot, because the others are virtually the same.
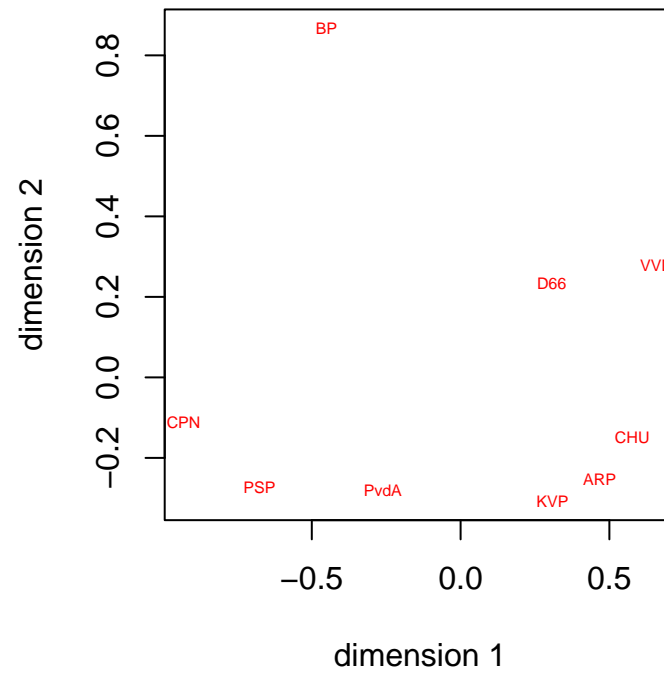
Gruijter, Ties = 1

We can also use this example to show the effect of using different initial configurations.

```
h1 <- smacofRO(gruijterData, 2, init = 1, labels = labels, verbose = FALSE)
h2 <- smacofRO(gruijterData, 2, init = 2, labels = labels, verbose = FALSE)
h3 <- smacofRO(gruijterData, 2, init = 3, labels = labels, verbose = FALSE)
```

- Init = 1 uses 88 iterations and stops at stress 0.0042180140
- Init = 2 uses 76 iterations and stops at stress 0.0039894695
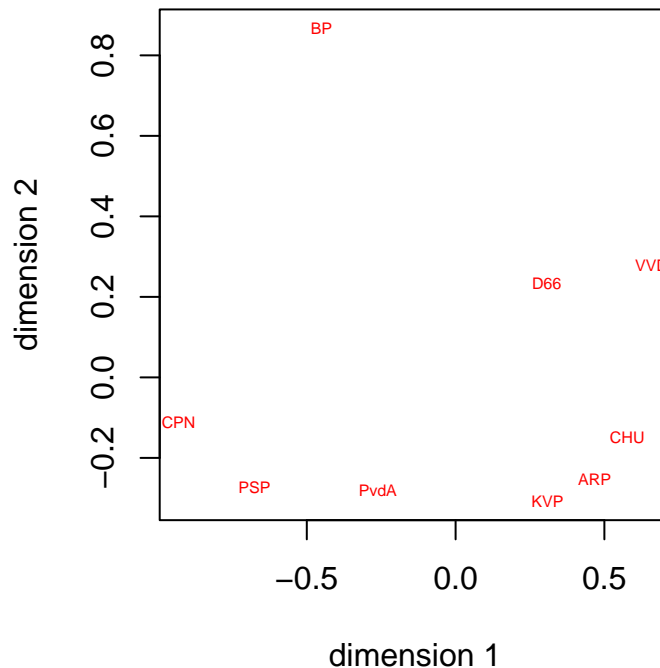- Init = 3 uses 69 iterations and stops at stress 0.0060410599

Gruijter with Torgerson Init

Gruijter with Maximum Sum Init
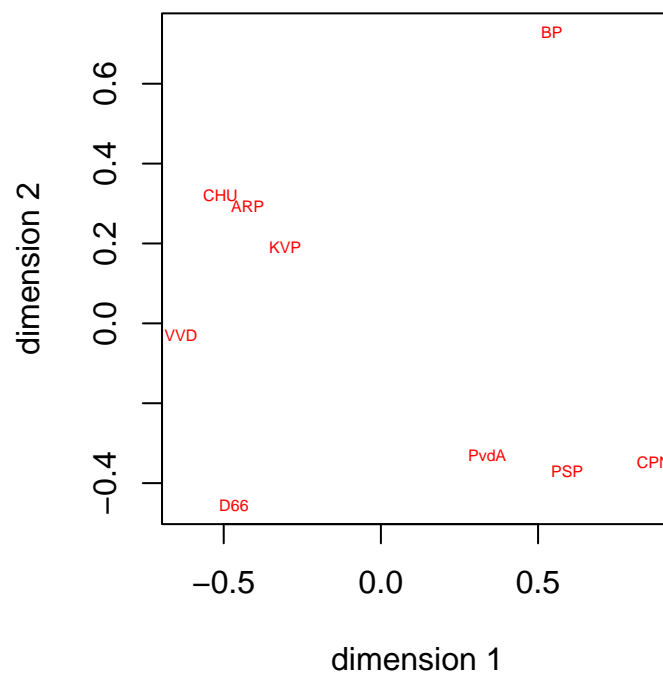
Gruijter with Random Init

Using different initial configurations in this example makes a huge difference. All three plots show the (CPN,PvdA,PSP) leftist cluster, the liberal (VVD,D'66) cluster, the protest BP outlier which is its own cluster, and the (KVP,ARP,CHU) christian democrat cluster. In the three plots the clusters are distributed differently over the plane.

4.2 Ekman (1954)

In the Ekman data there are 47 tieblocks out of 91 observations, and we expect the ties option to make some difference.

```
h1 <- smacofRO(ekmanData, 2, ties = 1, verbose = FALSE, labels = labels)
h2 <- smacofRO(ekmanData, 2, ties = 2, verbose = FALSE, labels = labels)
h3 <- smacofRO(ekmanData, 2, ties = 3, verbose = FALSE, labels = labels)
```
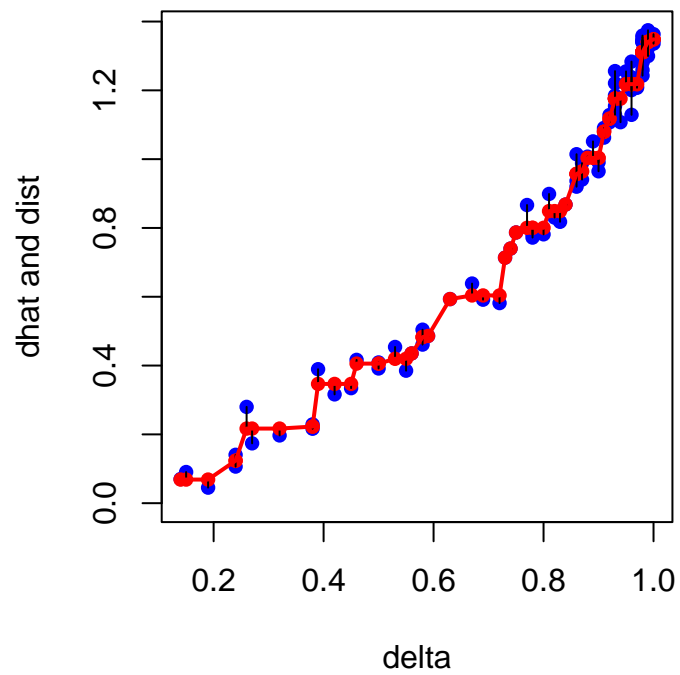
- Ties = 1 uses 40 iterations and stops at stress 0.0002668633
- Ties = 2 uses 32 iterations and stops at stress 0.0004988332
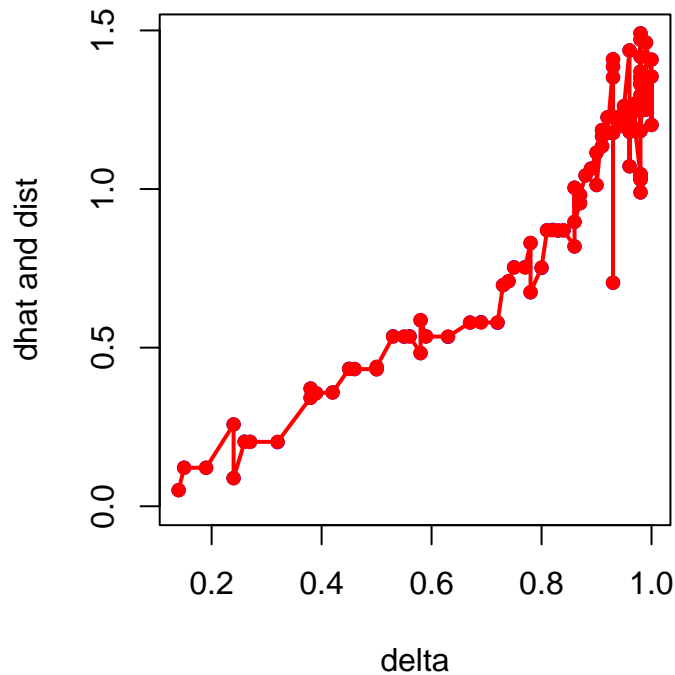- Ties = 3 uses 1000 iterations and stops at stress 0.0000001193

Ekman, Ties = 1
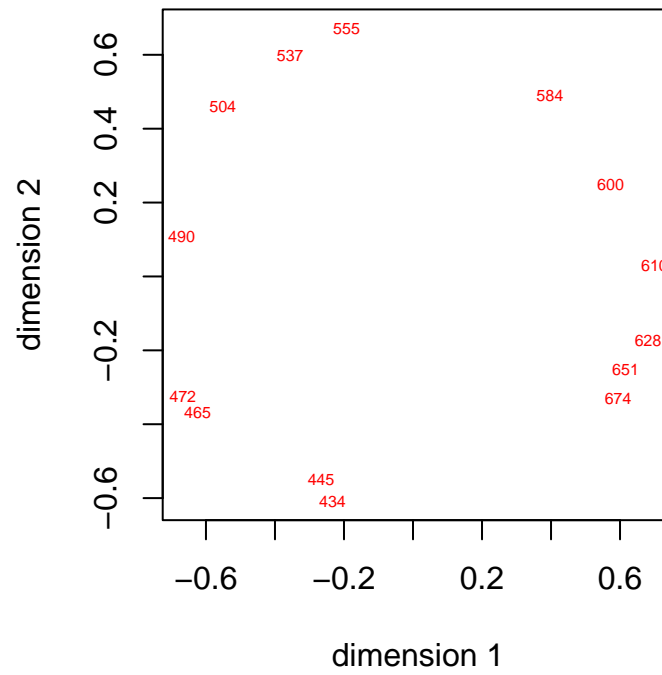

Ekman, Ties = 2

9
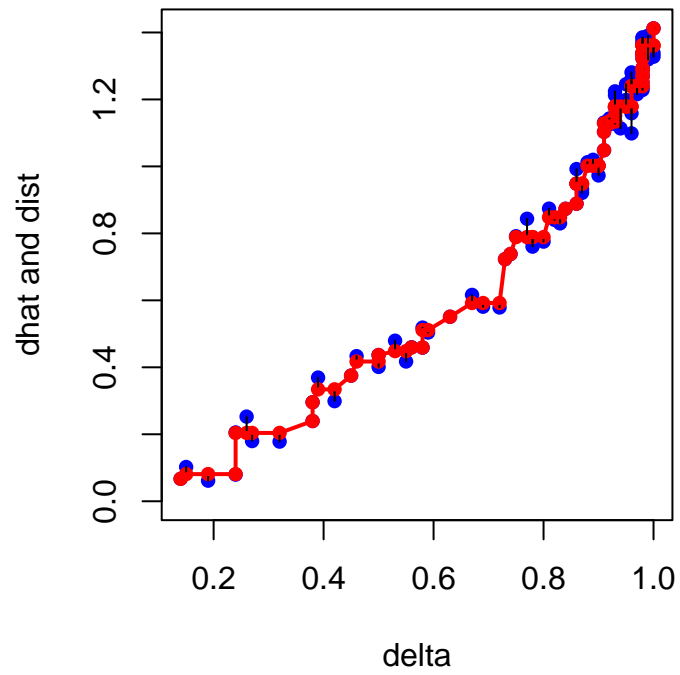
Ekman, Ties = 3

Although the stress values are indeed different, the solutions are practically the same. When looking at the Shepard plots we have to remember that for ties = 1 and ties = 3 we can have different disparity values for the same dissimilarity value. Thus tieblocks are represented as intervals on the vertical axis, and strictly spoken we do not have a functional relationship between delta and dhat. This is clear from the plot for the tertiary approach, which shows some of the intervals, mostly for the larger dissimilarities. It reinforces the idea that the tertiary approach is only useful if there are many small tieblocks, in which case it will be quite similar to the primary and secondary approach.
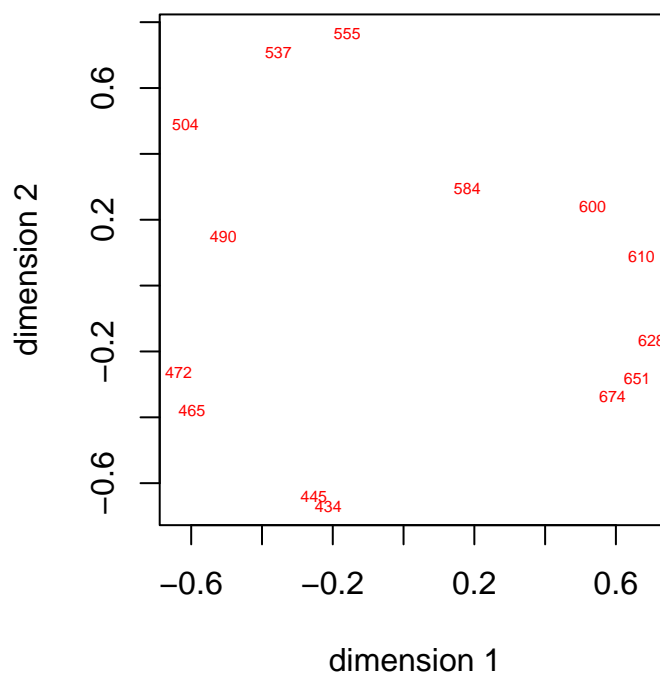
Ekman, Ties = 1

Ekman, Ties = 2
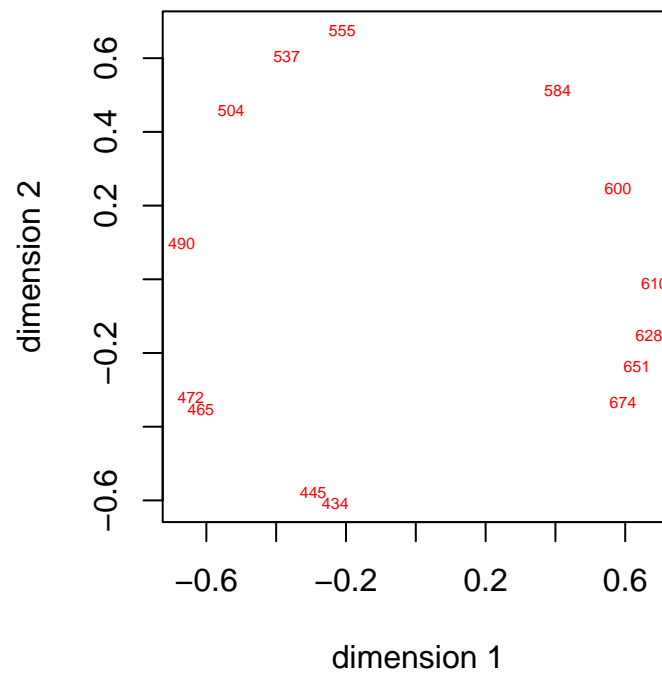
Ekman, Ties = 3



4.3 Rothkopf (1957)

```
h1 <- smacofRO(morseData, 2, ties = 1, verbose = FALSE, itmax = 10000)
h2 <- smacofRO(morseData, 2, ties = 2, verbose = FALSE, itmax = 10000)
h3 <- smacofRO(morseData, 2, ties = 3, verbose = FALSE, itmax = 10000)
```
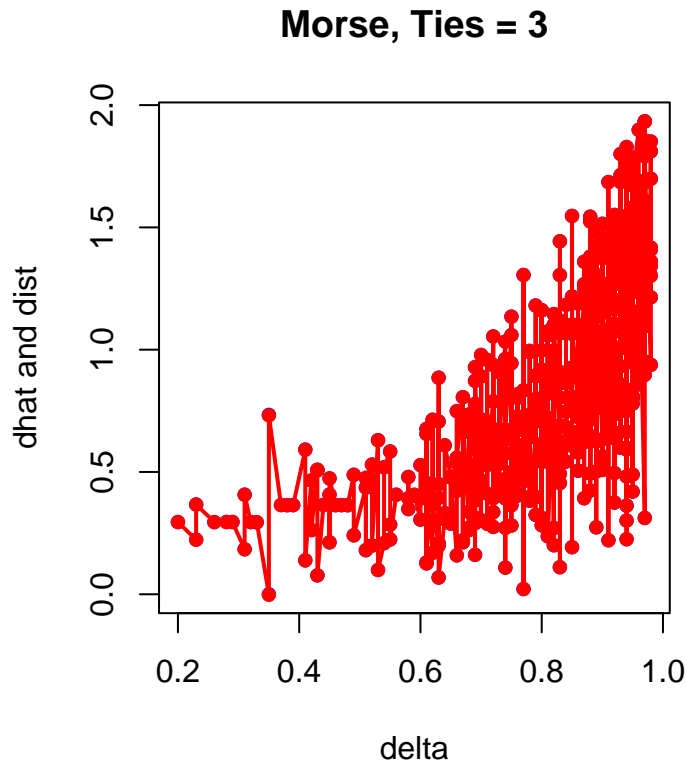
- Ties = 1 uses 34 iterations and stops at stress 0.0163278414
- Ties = 2 uses 30 iterations and stops at stress 0.0203202336
- Ties = 3 uses 1129 iterations and stops at stress 0.0000000226

Morse, Ties = 1
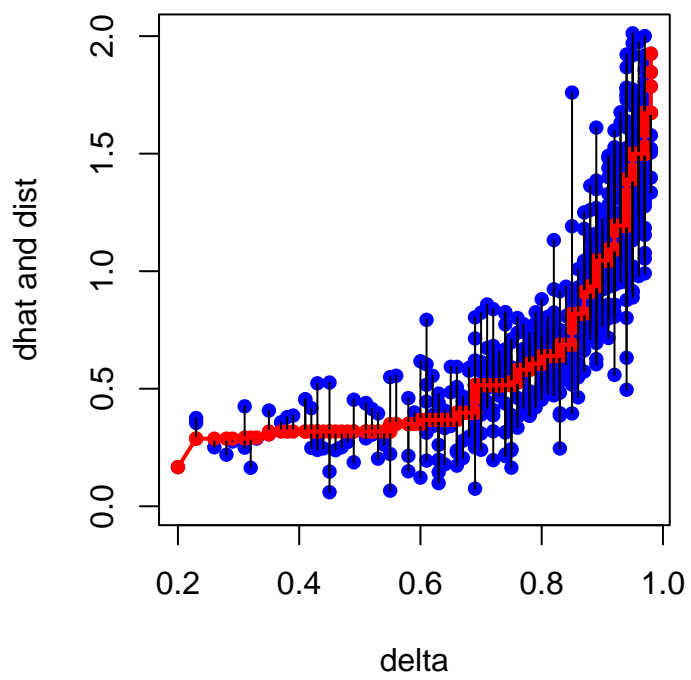
Morse, Ties = 2

Morse, Ties = 3
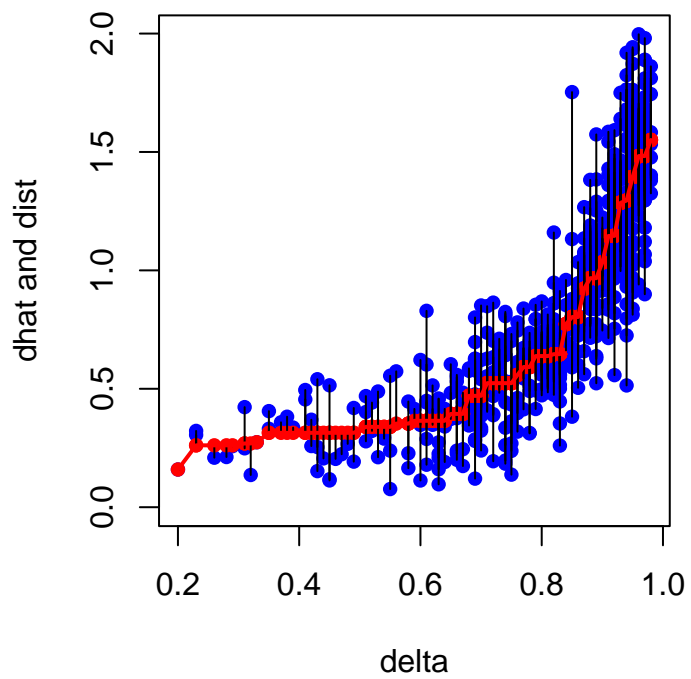
We also run morse with different values of kitmax, the number of inner Guttman iterations. The results are

```r
h1 <- smacofRO(morseData, 2, kitmax = 1, verbose = FALSE, itmax = 10000)
h2 <- smacofRO(morseData, 2, kitmax = 5, verbose = FALSE, itmax = 10000)
h3 <- smacofRO(morseData, 2, kitmax = 10, verbose = FALSE, itmax = 10000)
```

- Kitmax = 1 uses 138 major iterations and stops at stress 0.0163278417
- Kitmax = 5 uses 34 major iterations and stops at stress 0.0163278414
- Kitmax = 10 uses 22 major iterations and stops at stress 0.0163278414

We see that the resulting stress values are the same, but the number of outer iterations differs. This may be significant, because it means that using only one Guttman iteration means doing 138 monotone rgressions, while having 10 Guttman iterations per major iteration only uses 22 monotone regressions.

References

De Gruijter, D. N. M. 1967. "The Cognitive Structure of Dutch Political Parties in 1966." Report E019-67. Psychological Institute, University of Leiden.

De Leeuw, J. 1977. "Correctness of Kruskal's Algorithms for Monotone Regression with Ties." *Psychometrika* 42: 141–44.

———. 2017. "Exceedingly Simple Monotone Regression (with Ties)." 2017.

Ekman, G. 1954. "Dimensions of Color Vision." *Journal of Psychology* 38: 467–74.

Rothkopf, E. Z. 1957. "A Measure of Stimulus Similarity and Errors in some Paired-associate Learning." *Journal of Experimental Psychology* 53: 94–101.