Smacof at 50: A Manual
Part 2: Metric Smacof

Jan de Leeuw - University of California Los Angeles

Started December 12 2022, Version of May 03, 2024

Abstract

TBD

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

In this part of the manual we discuss metric MDS, and the program smacofAC.

2 Bells and Whistles

There are two options, *bounds* and *constant*, to make smacofAC more interesting and more widely applicable. Using these options the metric MDS problem becomes minimization of

$$\sigma(X, \hat{D}) = \sum\sum w_{ij}(\hat{d}_{ij} - d_{ij}(X))^2$$

over both X and $\hat{D}$, allowing some limited "metric" transformations of the data Δ. The four "metric" types of transformations are

1. type AC1: if bounds = 1 and constant = 1 $\delta^-_{ij} + c \leq \hat{d}_{ij} \leq \delta^+_{ij} + c$ for some c,
2. type AC2: if bounds = 0 and constant = 1 $\hat{d}_{ij} = \delta_{ij} + c$ for some c,
3. type AC3: if bounds = 1 and constant = 0 $\delta^-_{ij} \leq \hat{d}_{ij} \leq \delta^+_{ij}$,
4. type AC4: if bounds = 0 and constant = 0 $\hat{d}_{ij} = \delta_{ij}$.

In addition all four types require that $\hat{d}_{ij} \geq 0$ for all (i, j). Note that for types AC1 and AC2 the data Δ do not need to be non-negative. In fact, the original motivation for the additive constant in classical scaling (Messick and Abelson (1956)) was that Thurstonian analysis of paired or triadic comparisons produced dissimilarities on an interval scale, and thus could very well include negative values. The non-negativity requirement for $\hat{D}$ means in the case of AC2 that $c \geq -\min \delta_{ij}$. For AC1 it means $c \geq -\min \delta^+_{ij}$.

In AC1 and AC3 there is no mention of Δ, which means the bounds Δ^- and Δ^+ are actually the data. We could collect dissimilarity data by asking subjects for interval judgments. Instead of a rating scale with possible responses from one to ten we could ask for a mark on a line between zero and ten, and then interpret the marks as a choice of one of the intervals $[k, k+1]$. These finite precision or interval type of data could even come from physical measurements of distances. Thus the bounds parameter provides one way to incorporate uncertainty into MDS, similar to interval analysis, fuzzy computing, or soft computing.

makes sense to choose both Δ^- and Δ^+ to be monotone with Δ, although there is no requirement to do so.

2.1 Type AC2

Cooper (1972)

2.2 Type AC1

Of the four regression problems only the one with bounds = 1 and constant = 1 is non-trivial. It may help to give an example of what it actually requires. We use the De Gruijter example with $\delta^-_{ij} = \delta_{ij} - 1$ and $\delta^+_{ij} = \delta_{ij} - 1$.

```
##       KVP PvdA  VVD  ARP  CHU  CPN  PSP   BP
## PvdA 5.63
## VVD  5.27 6.72
```

```
## ARP  4.60 5.64 5.46
## CHU  4.80 6.22 4.97 3.20
## CPN  7.54 5.12 8.13 7.84 7.80
## PSP  6.73 4.59 7.55 6.73 7.08 4.08
## BP   7.18 7.22 6.90 7.28 6.96 6.34 6.88
## D66  6.17 5.47 4.67 6.13 6.04 7.42 6.36 7.36
```

The distances are taken from the Torgerson solution.

The Shepard plot is in figure 1. The two lines are connecting the δ_{ij}^- and the δ_{ij}^+, i.e. they give the bounds for $c = 0$. In our example the lines are parallel, because $\delta_{ij}^+ - \delta_{ij}^- = 2$ for all (i, j), but in general this may not be the case. We could, for example, set $\delta_{ij}^+ = (1+\alpha)\delta_{ij}$ and $\delta_{ij}^- = (1-\alpha)\delta_{ij}$, in which case the region between δ^- and δ^+ is like a trumpet. The bounds δ_{ij}^- and δ_{ij}^+ are not necessarily a monotone function of δ, in fact there may not even be a δ. Also note that the δ_{ij}^- are not necessarily non-negative.

The points between the two lines do not contribute to the loss, and the points outside the strip contribute by how much they are outside.

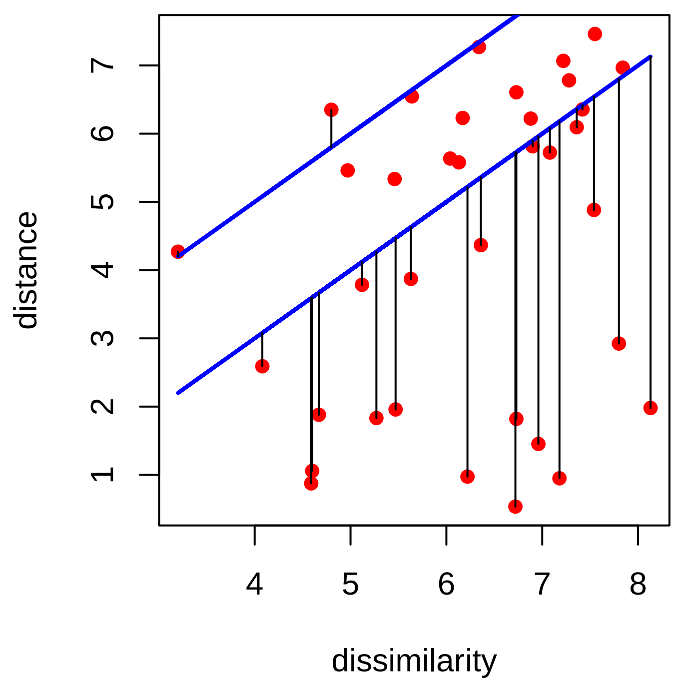


Figure 1: De Gruijter Shepard Plot

By varying c we shift the region between the two parallel lines upwards or downwards. The width of the region, or more generally the shape, always remains the same, because it is determined by the difference of δ^+ and δ^- and does not depend on c. The optimal c is the shift for which the red $(\delta_{ij}, d_{ij}(X))$ points are optimally within the strip between the δ^- and δ^+ lines. This is in the least squares sense, which means that we minimize the horizontal squared distances from the points outside the strip to the δ^- and δ^+ lines.

Let's formalize this. Define

$$\phi_{ij}(c) := \min_{\delta_{ij}\geq 0}\{(\delta_{ij}-d_{ij}(X))^2 \mid \delta_{ij}^- + c \leq \delta_{ij} \leq \delta_{ij}^+ + c\} \tag{1}$$

and

$$\phi(c) := \sum\sum w_{ij}\phi_{ij}(c) \tag{2}$$

The constraints are consistent if $\delta_{ij}^+ + c \geq 0$, i.e. if $c \geq c_0 := -\min \delta_{ij}^+$. The regression problem is to minimize ϕ over $c \geq c_0 := -\min \delta_{ij}^+$. At first sight it seems this we should require the stricter bound $c \geq -\min \delta_{ij}^-$, which implies that $\delta_{ij} \geq 0$ if $\delta_{ij}^- + c \leq \delta_{ij} \leq \delta_{ij}^+ + c$. If we use the more relaxed bound $c \geq c_0$ then $\delta_{ij}^- + c$ can be negative. But since $d_{ij}(X) \geq 0$ the algorithm never chooses a negative δ_{ij}, in fact $\delta_{ij} = 0$ if and only if $d_{ij}(X) = 0$.

Figure 2 has an example of one of the ϕ_{ij}. The value of $d_{ij}(X)$ is 4.2698418, δ_{ij}^0 is 3.2, δ_{ij}^- is 2.2, and δ_{ij}^+ is 4.2. The two vertical lines are at $c = d_{ij}(X) - \delta_{ij}^+$ and $c = d_{ij}(X) - \delta_{ij}^-$. Between those two lines ϕ_{ij} is zero bercause $\hat{d}_{ij} = d_{ij}(X)$. If $c \geq d_{ij}(X) - \delta_{ij}^-$ then $\hat{d}_{ij} = \delta_{ij}^- + c$ and ϕ_{ij} is the quadratic $(d_{ij}(X) - (\delta_{ij}^- + c))^2$. If $c \leq d_{ij}(X) - \delta_{ij}^+$ then $\hat{d}_{ij} = \delta_{ij}^+ + c$ and ϕ_{ij} is the quadratic $(d_{ij}(X) - (\delta_{ij}^+ + c))^2$. It follows that ϕ_{ij} is piecewise quadratic, convex, and continuously differentiable. The derivative is piecewise linear, continuous, and increasing. In fact

$$\mathcal{D}\phi_{ij}(c) = \begin{cases} 2(c-(d_{ij}(X)-\delta_{ij}^+)) & \text{if } c \leq d_{ij}(X)-\delta_{ij}^+, \\ 2(c-(d_{ij}(X)-\delta_{ij}^-)) & \text{if } c \geq d_{ij}(X)-\delta_{ij}^- \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

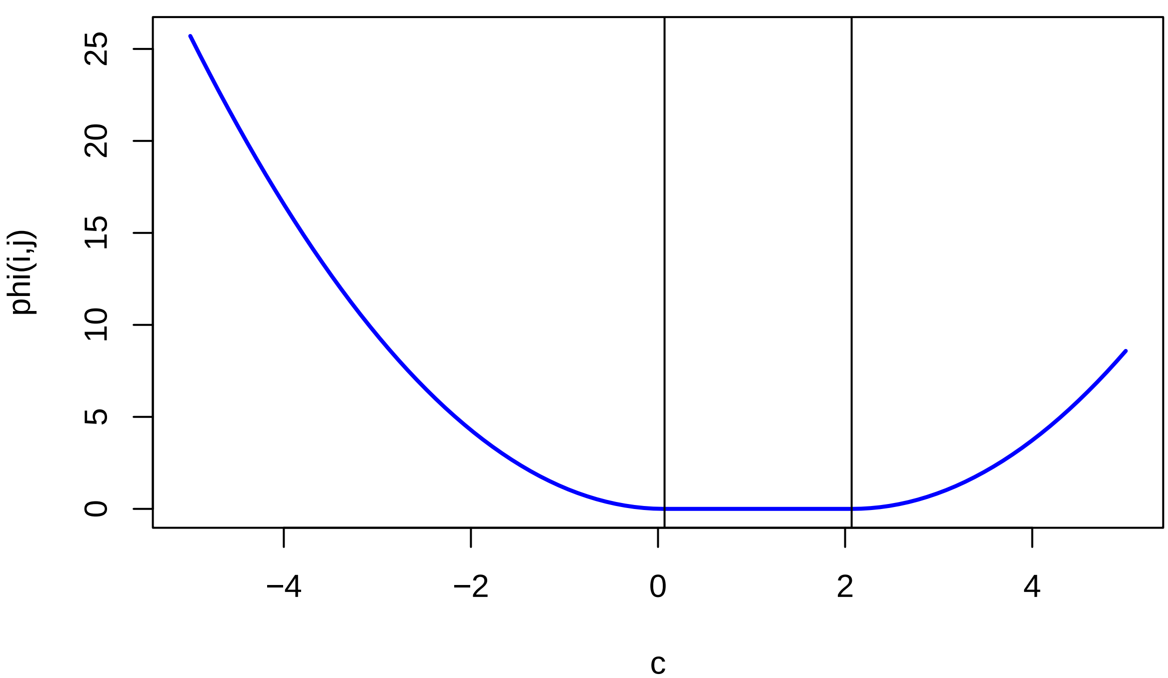


Figure 2: De Gruijter Example, One Phi_ij

Since ϕ is a positive linear combination of the ϕ_{ij} it is also piecewise quadratic, convex, and continuously differentiable with a continuous piecewise linear increasing derivative. Note ϕ is **not** twice-differentiable and **not** strictly convex. Figure 3 has a plot of ϕ for the De Gruijter example. The red vertical line is at $c = c_0$.

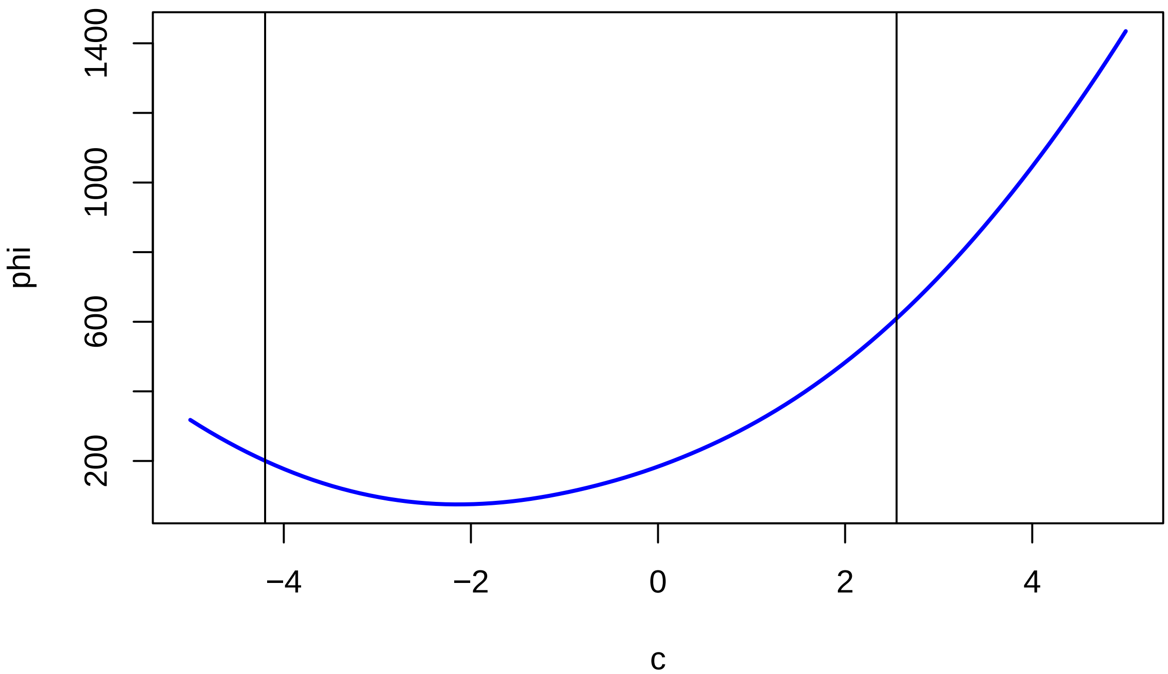


Figure 3: Function Phi for De Gruijter Example

We minimize ϕ by using the R function optimize(). From (3) we see that if $c \geq \max d_{ij}(X) - \delta_{ij}^{-}$ then $\mathcal{D}\phi(c) \geq 0$. This can be used for the right endpoint of the interval over which we minimize ϕ, with c_0 as the left endpoint. The minimum of ϕ in our example turns out to be 75.251647, attained at c equal to -2.1403321. The Shepard plot corresponing with the optimum c is plotted in figure 4.

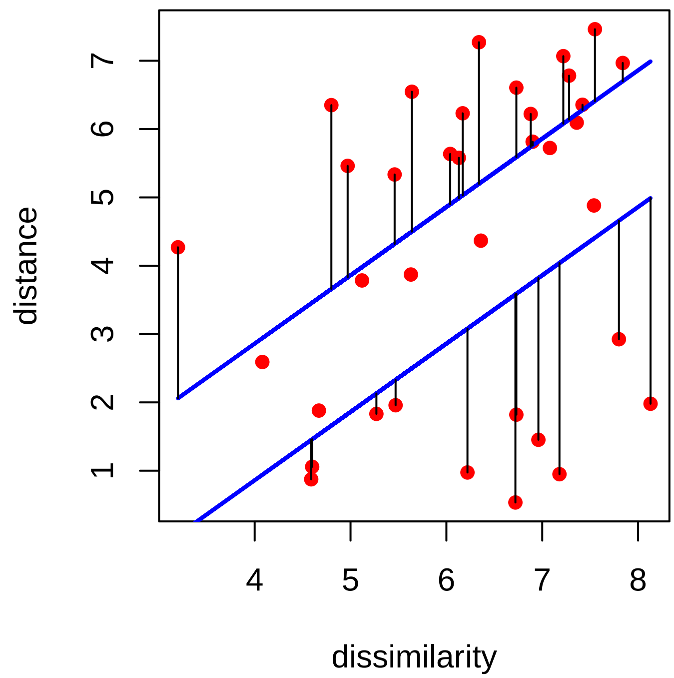
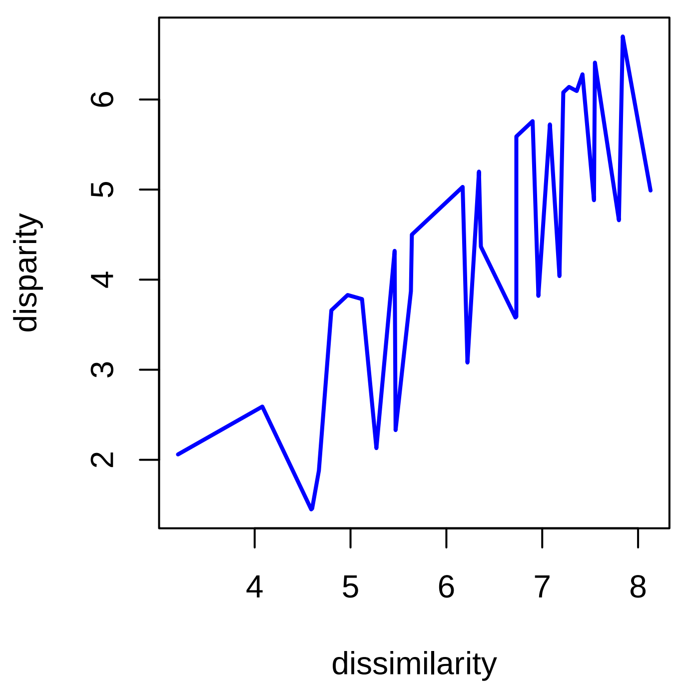
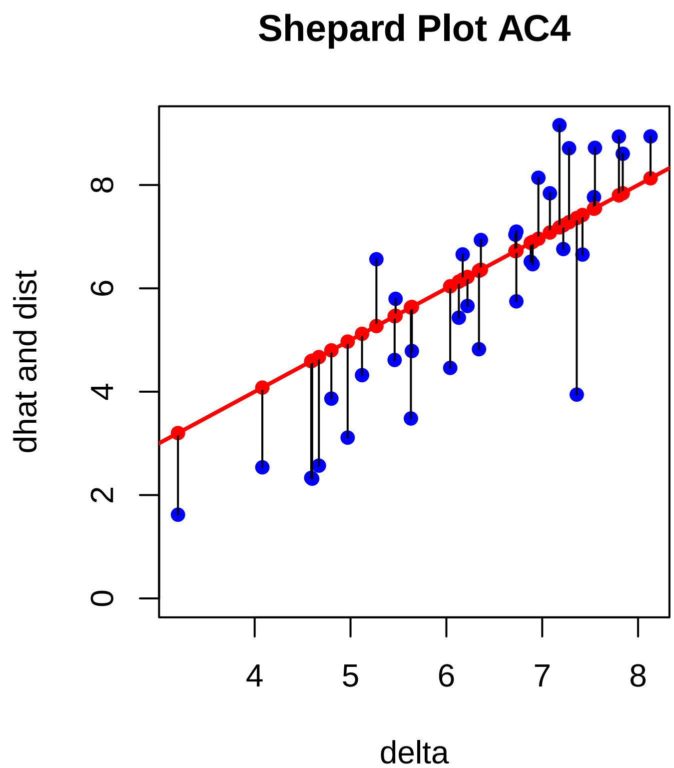


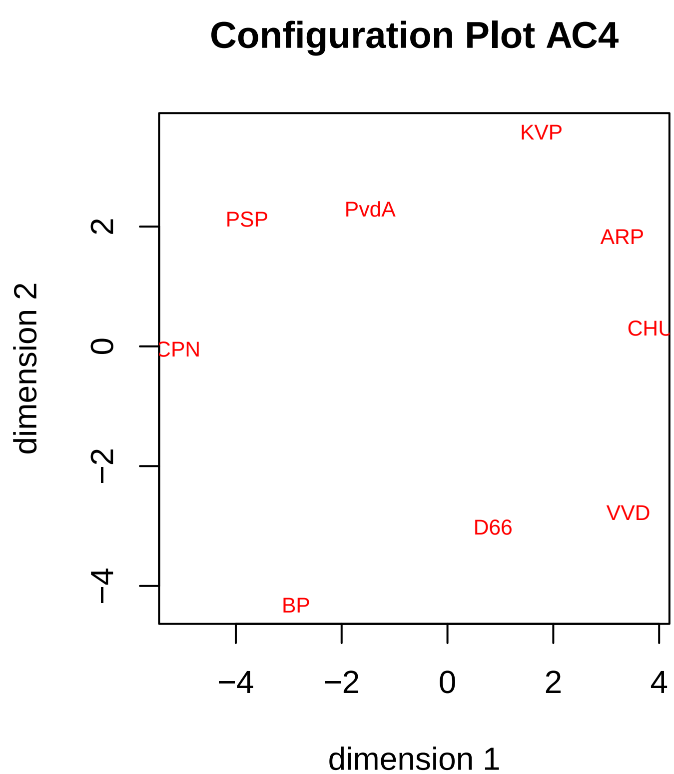
Figure 4: De Gruijter Shepard Plot

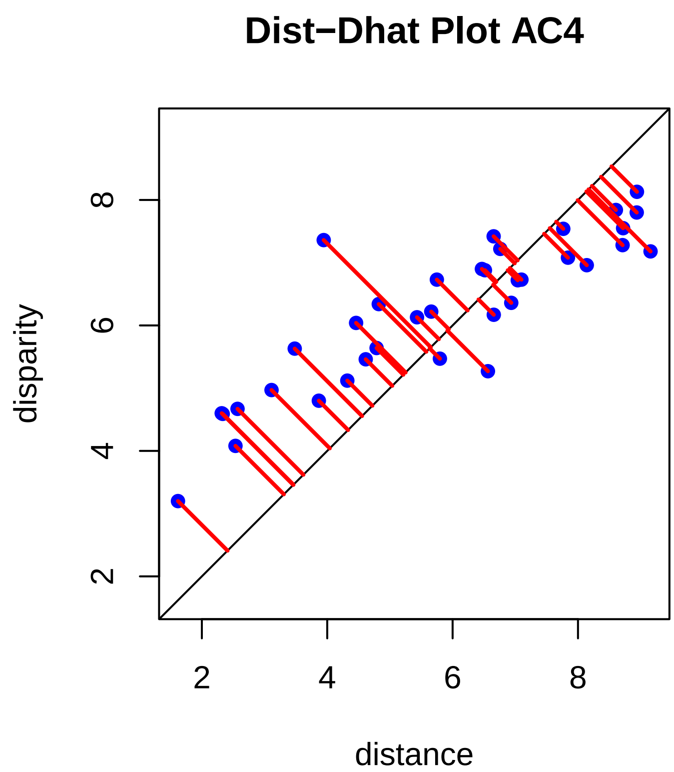


3 Example

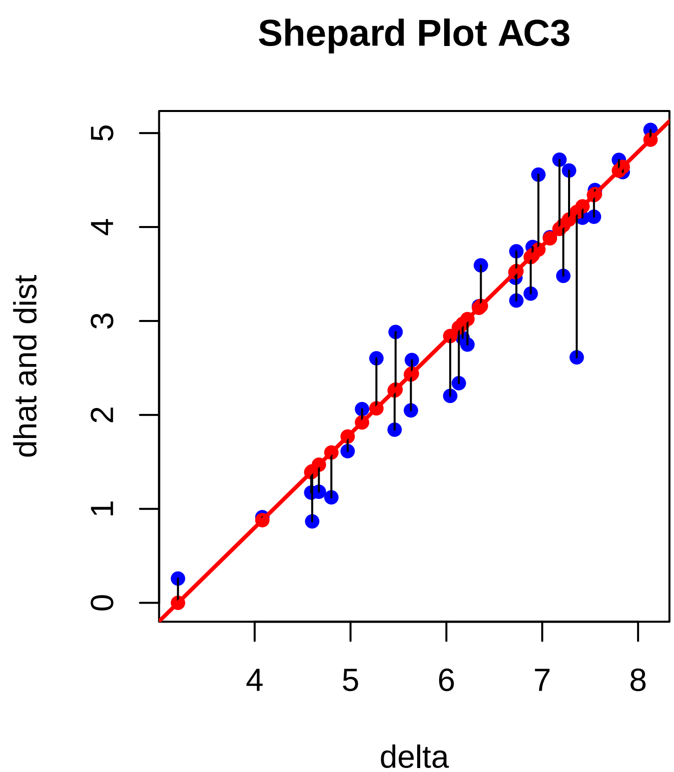
3.1 Type AC4

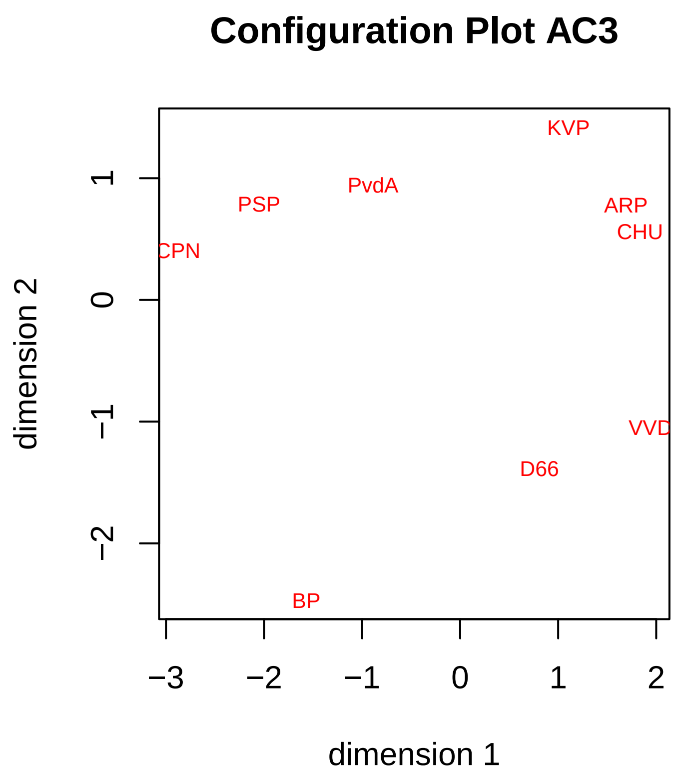


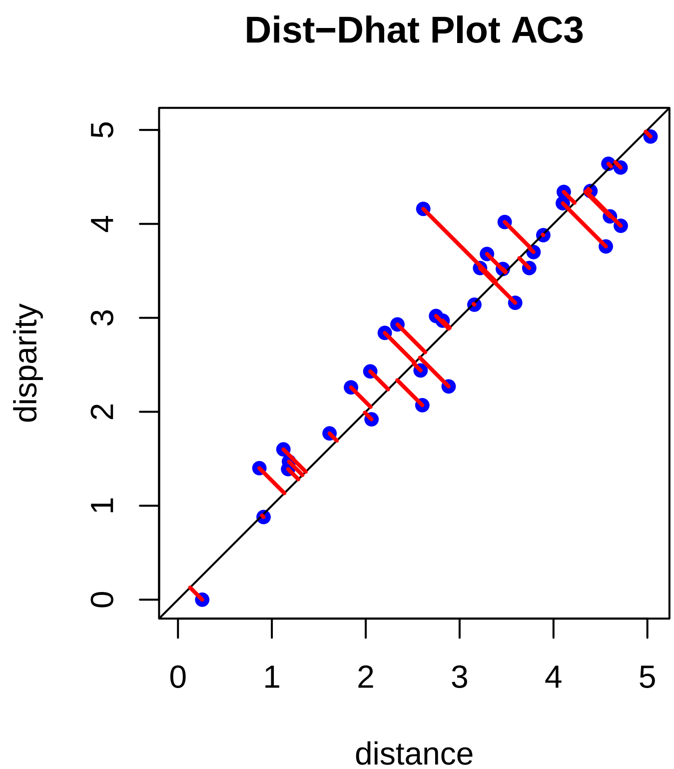



Dist−Dhat Plot AC4
disparity
distance
2
4
6
8

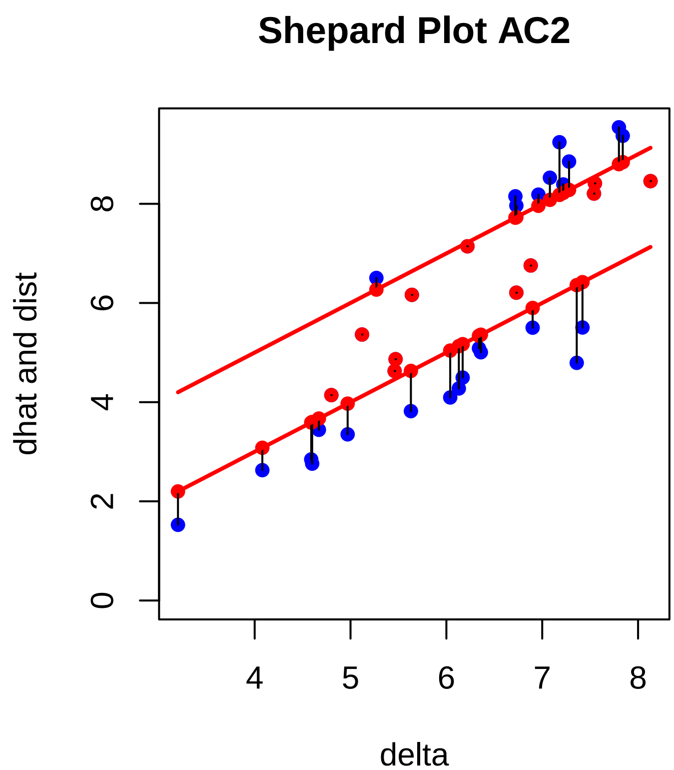
3.2 Type AC3

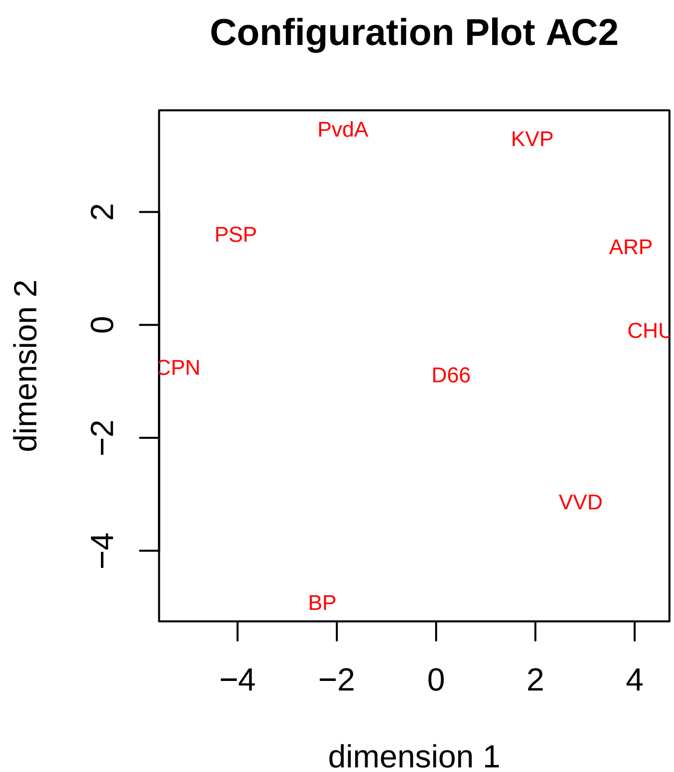

Shepard Plot AC3
dhat and dist
delta
0
1
2
3
4
5
6
7
8

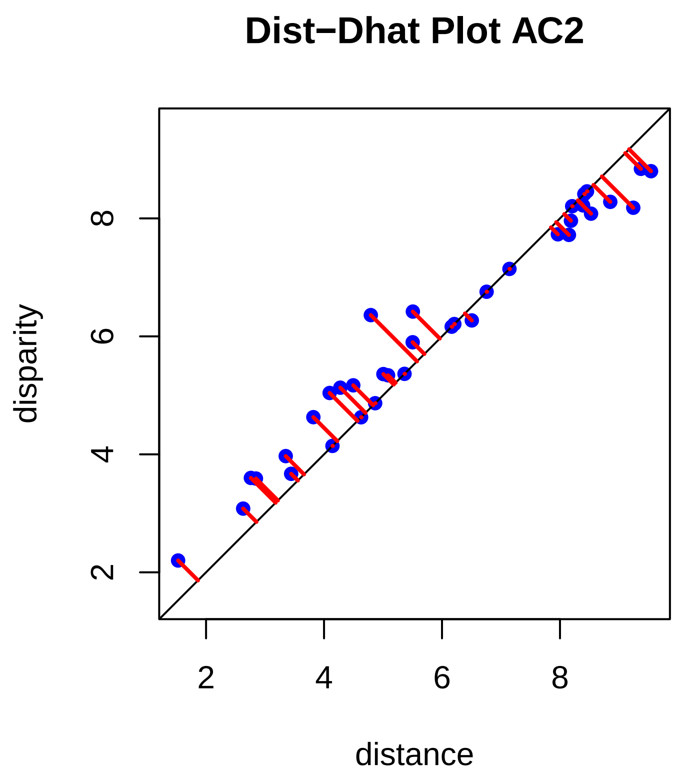

Configuration Plot AC3
KVP
PvdA
PSP
ARP
CHU
CPN
VVD
D66
BP
dimension 2
dimension 1
1
0
−1
−2
−3
−2
−1
0
1
2


Dist−Dhat Plot AC3
disparity
distance
0
1
2
3
4
5
0
1
2
3
4
5

3.3 Type AC2

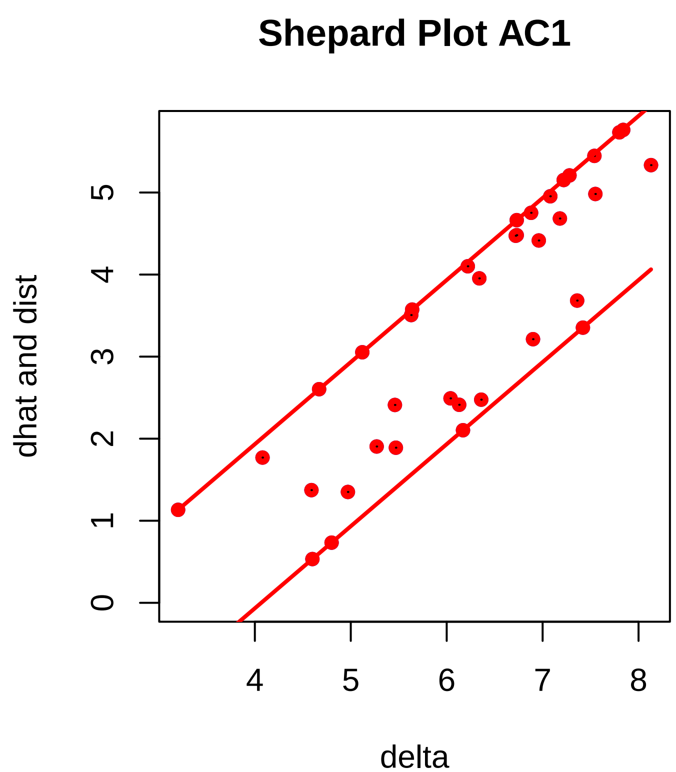

Shepard Plot AC2
dhat and dist
0
2
4
6
8
4
5
6
7
8
delta


Configuration Plot AC2
PvdA
KVP
PSP
ARP
CHU
CPN
D66
VVD
BP
dimension 2
−4
−2
0
2
−4
−2
0
2
4
dimension 1

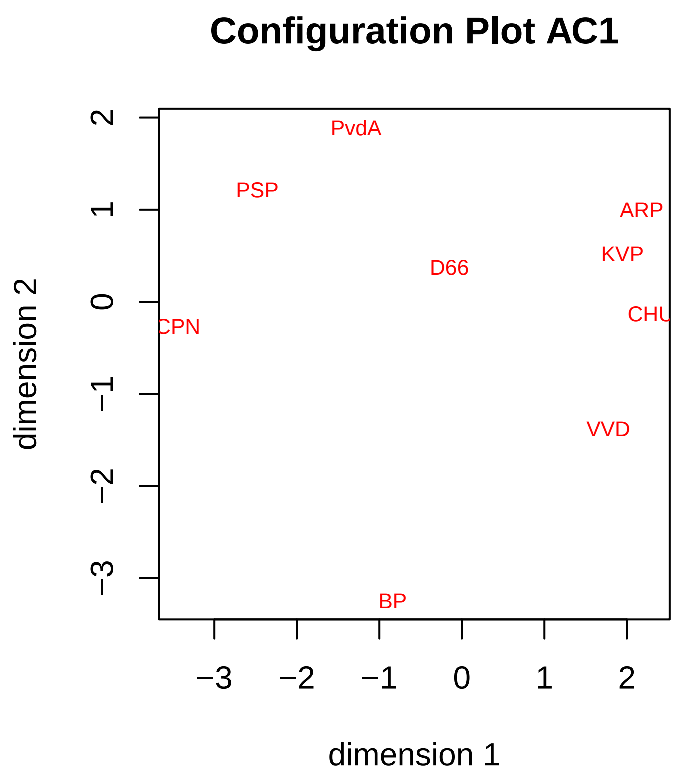


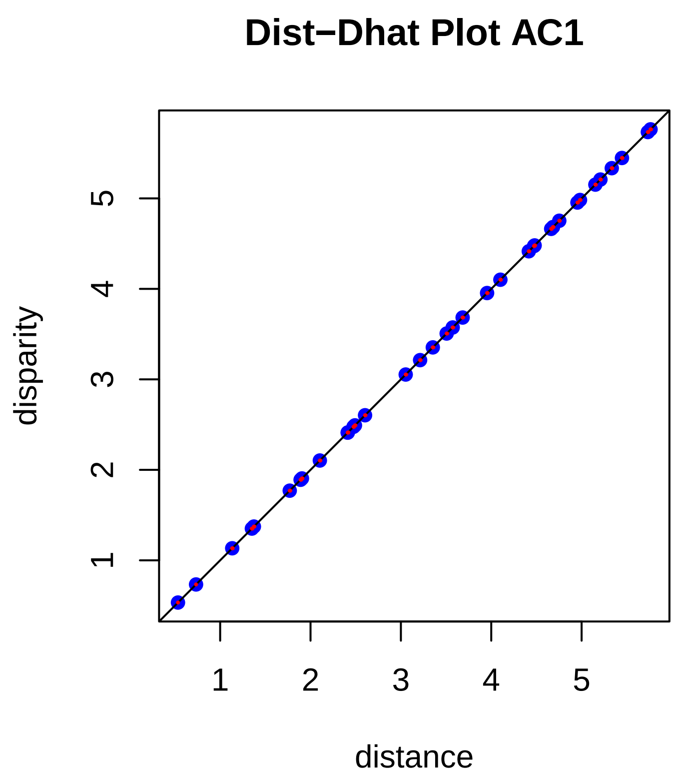
Dist−Dhat Plot AC2
disparity
distance
2
4
6
8

3.4 Type AC1



Shepard Plot AC1
dhat and dist
delta
0
1
2
3
4
5
6
7
8


Configuration Plot AC1
PvdA
PSP
ARP
KVP
D66
CHU
CPN
VVD
BP
dimension 2
dimension 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2


Dist−Dhat Plot AC1
disparity
distance
5
4
3
2
1
1
2
3
4
5

References

Cooper, L. G. 1972. "A New Solution to the Additive Constant Problem in Metric Multidimensional Scaling." *Psychometrika* 37 (3): 311–22.

Messick, S. J., and R. P. Abelson. 1956. "The Additive Constant Problem in Multidimensional Scaling." *Psychometrika* 21 (1–17).