Tertiary Approach Considered Harmful

Jan de Leeuw

November 23, 2025

In ordinal non-metric scaling Kruskal (1964a) introduced two different ways to handle ties in the data. De Leeuw (1977) added a third one, appropriately called the tertiary approach. In this paper we point out some problems that can occur when using the tertiary approach and interpreting its results.

Table of contents

Note: This is a working manuscript which may be expanded/updated multiple times. 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/tertiary

1 Introduction

In the monotone regression problem considered in this paper the data consist of the *target*, a numerical vector y of length n, and a partitioning of $\mathcal{N} := \{1, 2, \cdots, n\}$ into $m \leq n$ subsets $\mathcal{N}_k$, called *tie-blocks*. Tie-blocks $\mathcal{N}_k$ is *earlier* than tie-block $\mathcal{N}_\ell$ if and only if $k < \ell$. We also construct a total partial order $\preceq$ on $\mathcal{N}$, with corresponding $i \equiv j$ if both $i \preceq j$ and $j \preceq i$, and $i \prec j$ if $i \preceq j$ and not $j \preceq i$. Define $i \equiv j$ if i and j are in the same tie-block and $i \prec j$ if i is in an earlier tie-block than j.

In the least squares version of monotone regression in which all tie-blocks have only a single element we minimize the weighted least squares loss function

$$\sigma(x) := \sum_{i=1}^{n} w_i (x_i - y_i)^2 \tag{1}$$

over all numerical x with $x_1 \leq x_2 \leq \cdots x_n$. This special case of the constraints, which leads to a simple $O(n)$ algorithm, is called the *linear case*.

If there are ties then the user of a non-metric scaling programm typically have to choose from various options. The two most prominent ones, proposed and baptized by Kruskal (1964a) and Kruskal (1964b), are the *primary* and *secondary* approach to ties.

In the primary approach we require $x_i \leq x_j$ if $i \prec j$ and there are no constraints within tie-blocks. Kruskal (1964b) showed that primary approach monotone regression can be solved by requiring $x_i \leq x_j$ if $i \prec j$ but also if $i \approx j$ and $y_i < y_j$. Thus y is used to order the indices within tie-blocks. If $i \approx j$ as well as $y_i = y_j$ we complete the total order by requiring either $x_i \leq x_j$ or $x_j \leq x_i$. Which one of the two choices we make does not matter for the outcome of the monotone regression. There are no more ties and we proceed as in the linear case. A formal proof that this solves the primary case is in De Leeuw (1977).

In the secondary approach we require $x_i \leq x_j$ if $i \preceq j$, which implies $x_i = x_j$ if $i \equiv j$. Kruskal (1964b) also that the secondary approach can be solved by a linear case algorithm on the tie-block weighted averages

$$\overline{y}_k = \frac{\sum_{i \in \mathcal{N}_k} w_i y_i}{\sum_{i \in \mathcal{N}_k} w_i} \tag{2}$$

with tie-block weights

$$\overline{w}_k = \sum_{i \in \mathcal{N}_k} w_i. \tag{3}$$

For this monotone regression the solution is, say, $\overline{x}_k$. The secondary approach then has the solution x with x_i equal to the $\overline{x}_k$ of its tie-block. For a formal proof, again see De Leeuw (1977).

2 The Tertiary Approach

The tie-block averages $\overline{y}_k$, and their monotone regression $\overline{x}_k$, were also used by De Leeuw (1977) to define a *tertiary* approach, which requires $\overline{x}_k \leq \overline{x}_\ell$ if $k < \ell$. There are no other constraints on the x_i. It is shown in De Leeuw (1977) that the solution for x is given by

$$x_i = \overline{x}_k + (y_i - \overline{y}_k). \tag{4}$$

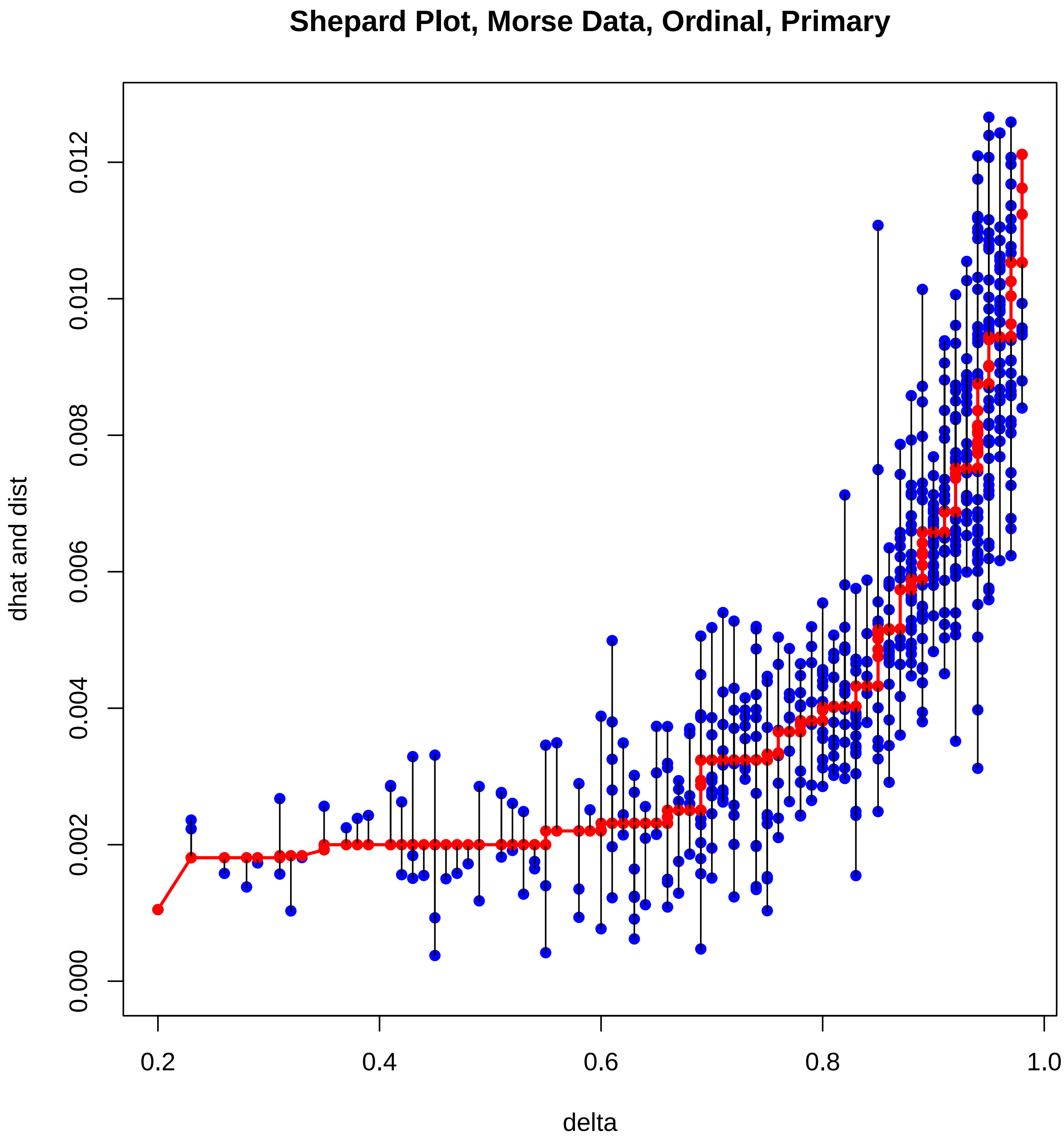
All three approaches are implemented as options in the smacof package (De Leeuw and Mair (2009), Mair, Groenen, and De Leeuw (2022)) and the smacofx package (Rusch et al. (2025)).

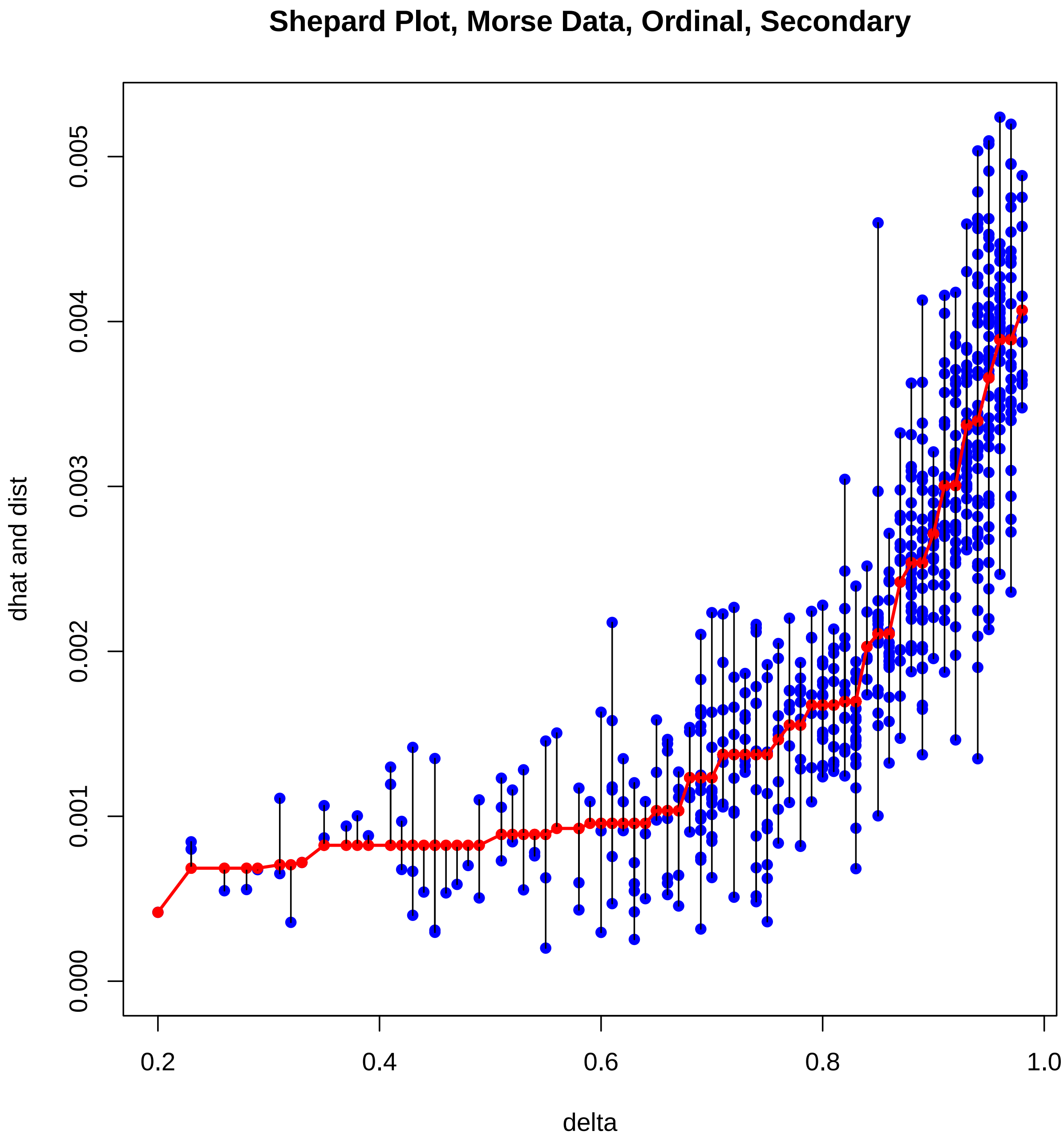
If there are only a few ties and there is a good fit the difference between the three approaches will be small. But there are several problems with the tertiary approach, and users of the smacof program should think twice before using it.

Before pointing out what the problems are we first analyze an example. We use the symmetrized version of the morse code data of Rothkopf (1957), available in the smacof package. The morse code data are analyzed three times, with the primary, secondary, and tertiary approach. We iterate until stress, the MDS loss function, changes less than 10^{-10} from one iteration to the next. All weights are one.

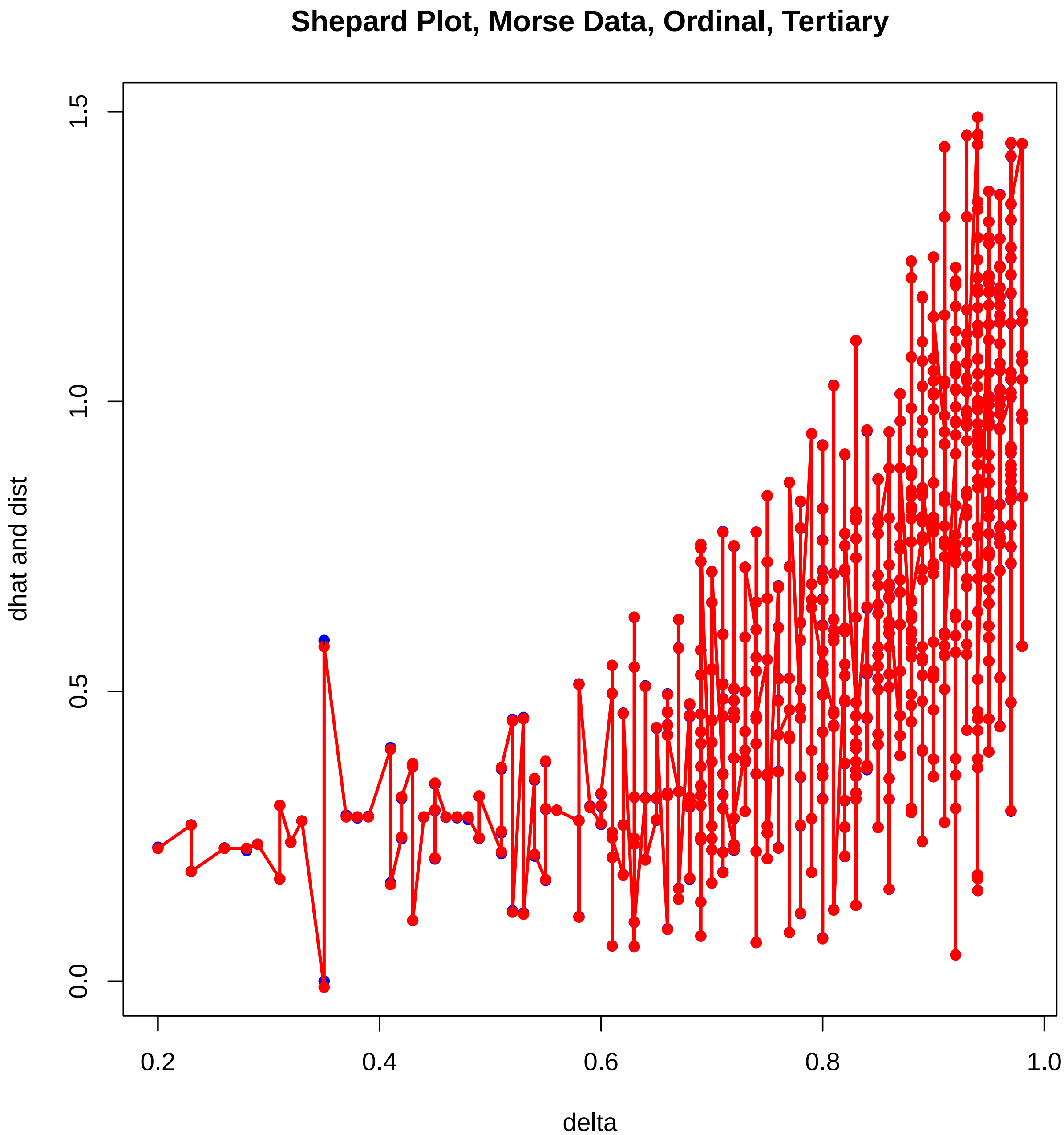
For the primary approach the analysis takes 143 iterations, with final loss function value 0.0337580759. For the secondary approach this is 136 iterations and loss 0.0423620820. For the tertiary approach we use 351 iterations and find loss 0.0000018073.

For each analysis we give two plots. The first three plots are the Shepard plots, with dissimilarities on the horizontal axis and final distances and disparities (i.e. the final monotone regression of the distances) on the vertical axis. The next three plots have final distance on the horizontal axis and disparity on the vertical axis.

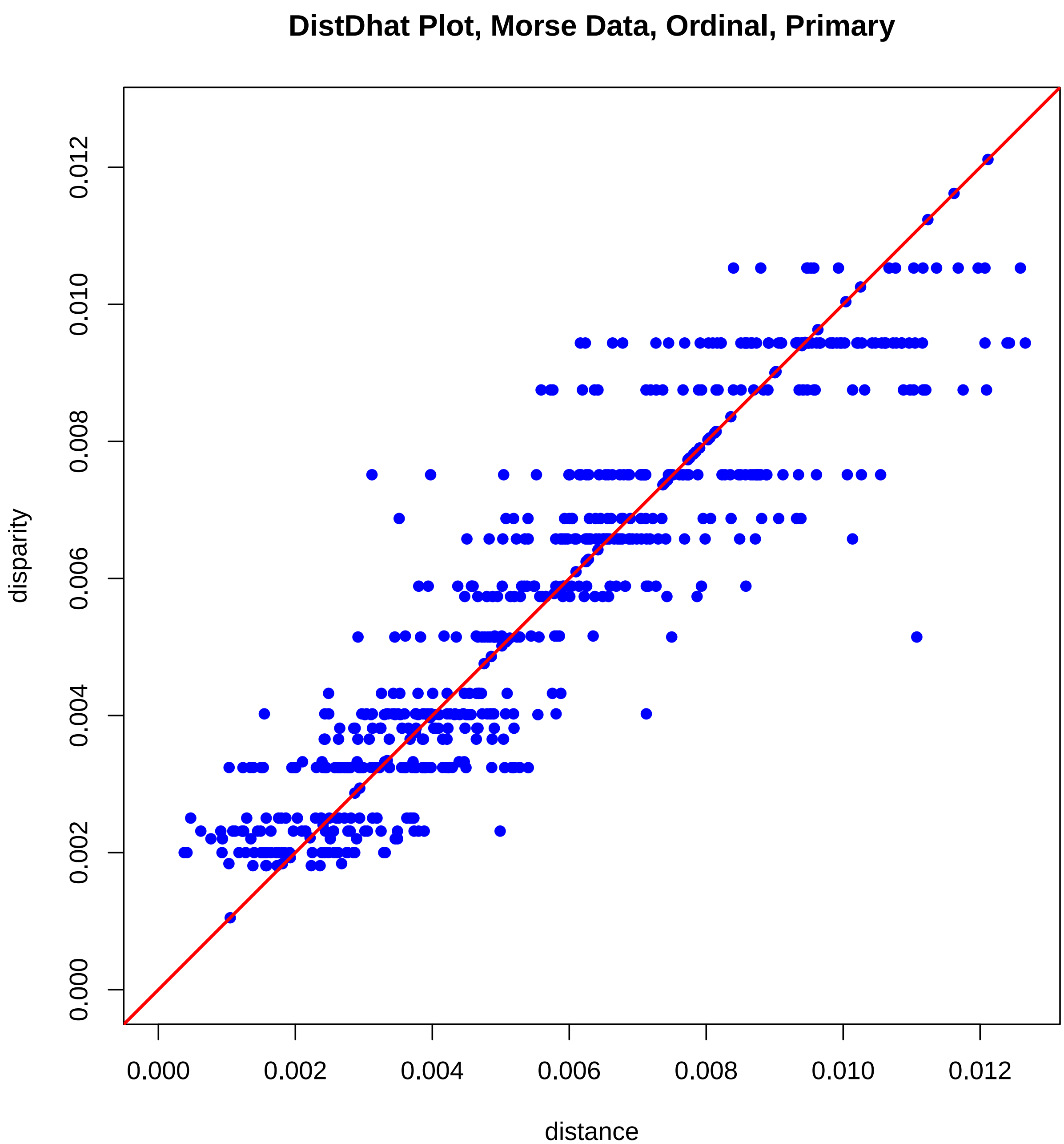

Shepard Plot, Morse Data, Ordinal, Primary
dhat and dist
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.2
0.4
0.6
0.8
1.0
delta



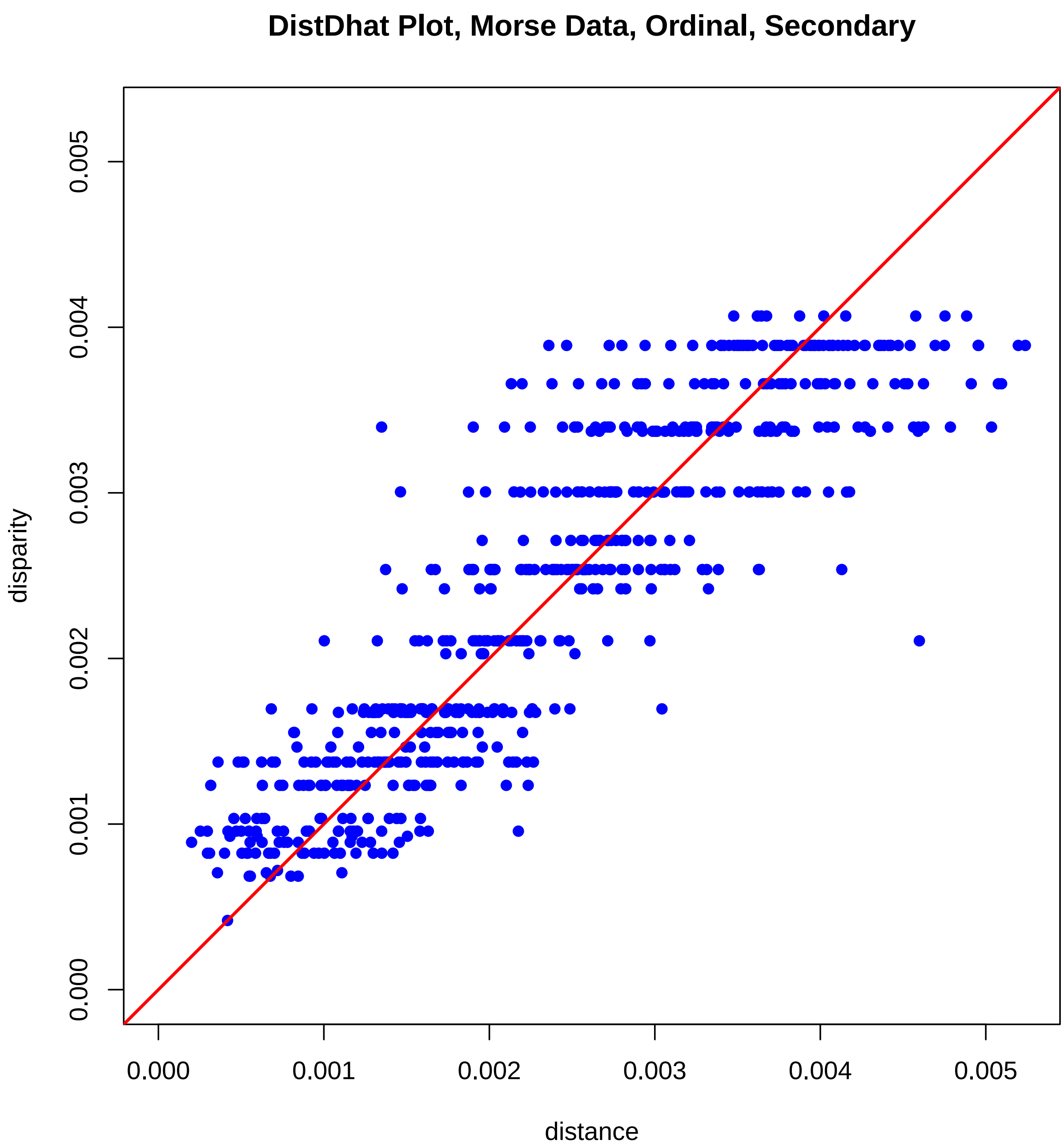
Shepard Plot, Morse Data, Ordinal, Secondary
dhat and dist
0.000
0.001
0.002
0.003
0.004
0.005
0.2
0.4
0.6
0.8
1.0
delta



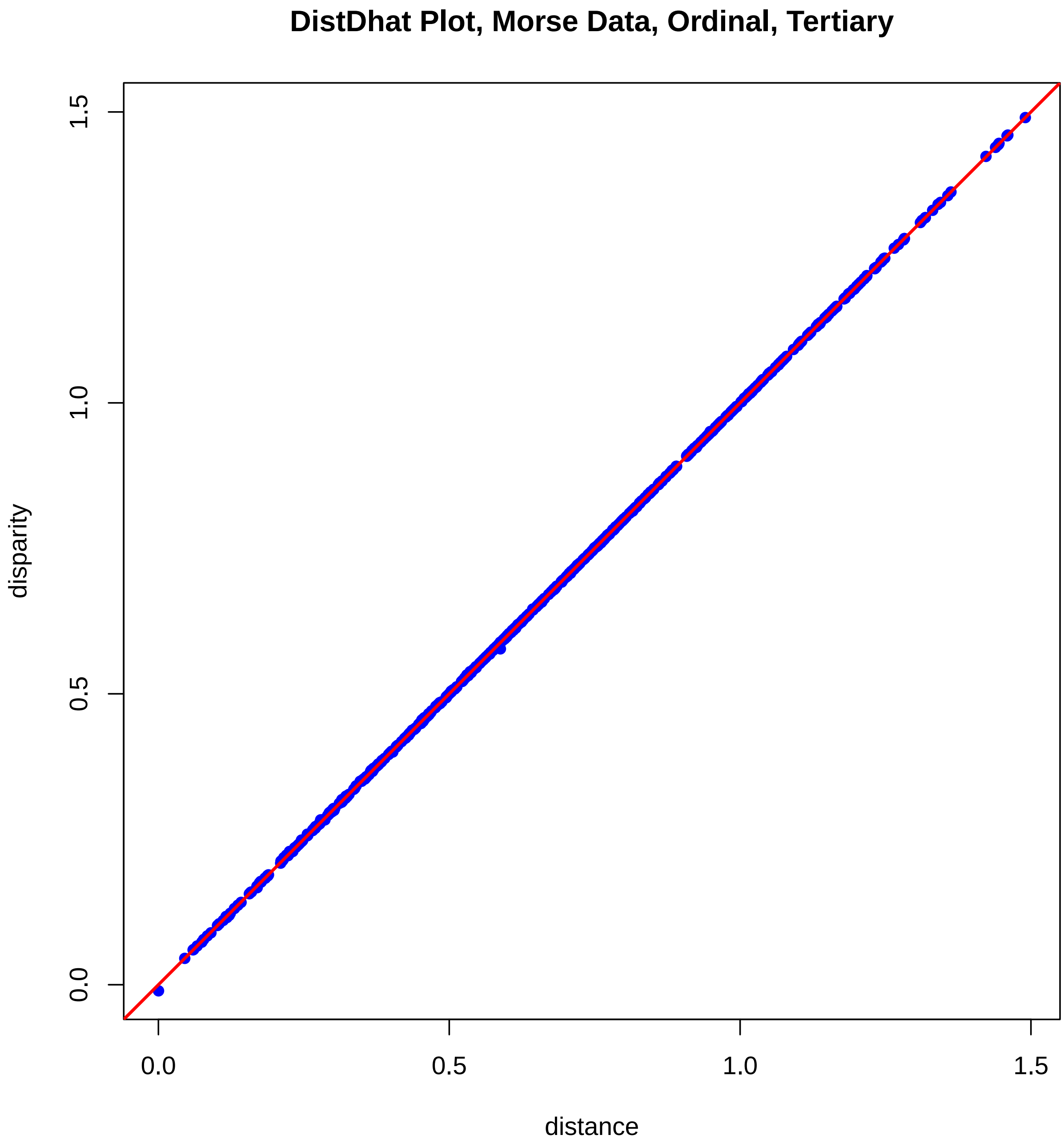
Shepard Plot, Morse Data, Ordinal, Tertiary
dhat and dist
0.0
0.5
1.0
1.5
0.2
0.4
0.6
0.8
1.0
delta



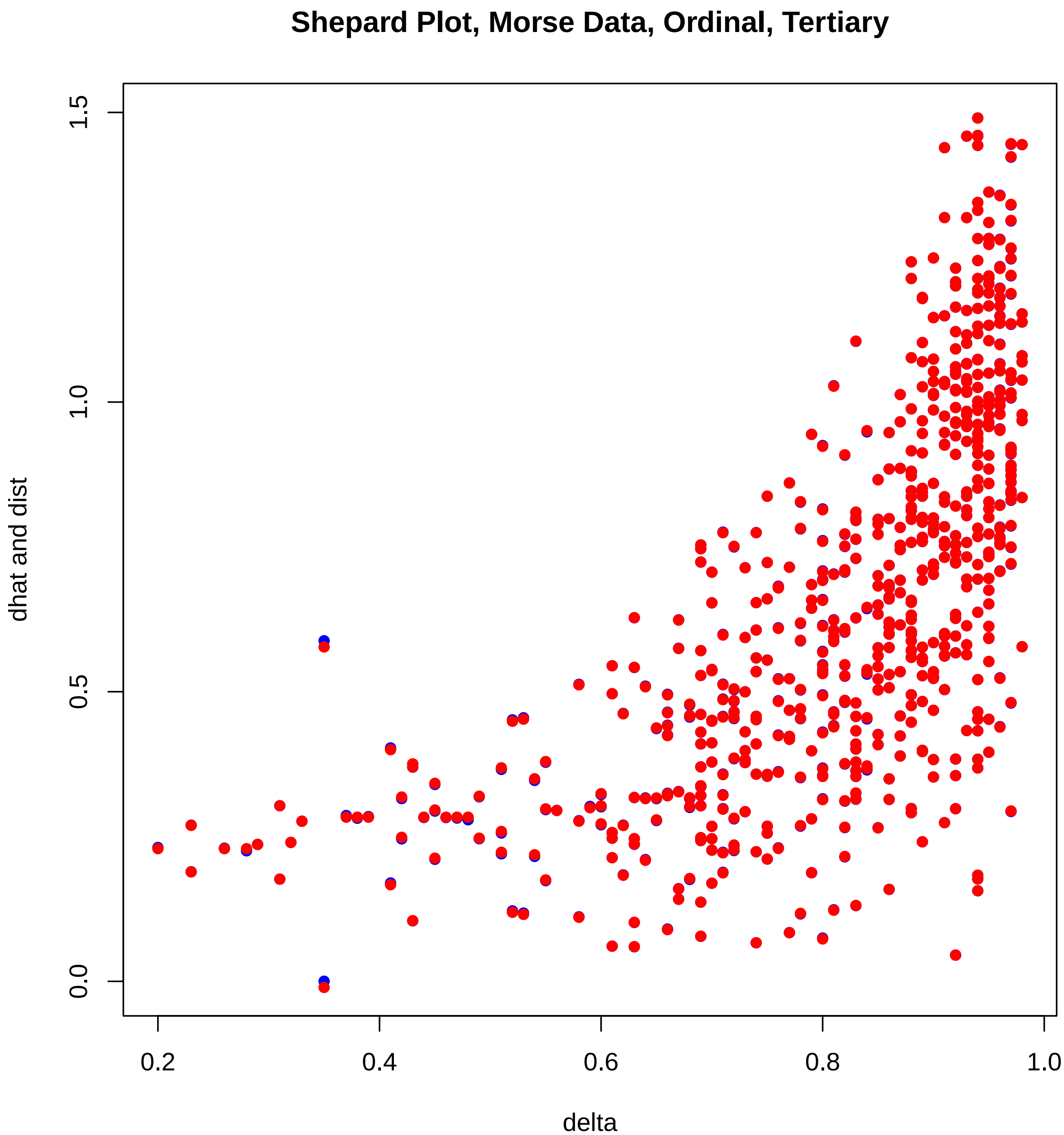
DistDhat Plot, Morse Data, Ordinal, Primary
disparity
0.000
0.002
0.004
0.006
0.008
0.010
0.012
distance
0.000
0.002
0.004
0.006
0.008
0.010
0.012



DistDhat Plot, Morse Data, Ordinal, Secondary
disparity
0.000
0.001
0.002
0.003
0.004
0.005
distance



The plots illlustrate some of the problems with the tertiary approach. In examples like this, with a fairly large numbers of ties and a rather bad fit using the primary or secondary approach, the Shepard plots for the tertiary approach are very far from monotone. The monotone trend may still be there but the plot goes up and down when connecting successive points. To some extend this is emphasized by drawing the connection lines, but even if we leave them out the scatterplot looks like a bad fit.



This brings us to the next problem with the tertiary approach. Although the Shepard plot suggests a bad fit, the dist-dhat plot shows an almost perfect fit, corresponding with the very small value of the final stress. There is nothing really wrong with this result, but it can be misleading. The fit of the tertiary approach may not be too good to be true, but it may be too good to be useful.

There is a more serious problem that can happen, although it does not occur in the morse code example. Consider a small example with two tie-blocks, each having an equal number of elements. All weights are one. If the tie-block means $\overline{y}_1$ and $\overline{y}_2$ are in the correct order, then monotone regression with the tertiary approach gives $x_i = y_i$ and we have perfect fit. If $\overline{y}_1 > \overline{y}_2$ then $\overline{x}_1$ and $\overline{x}_2$ are both equal to $\overline{y}$, the overall mean. Thus for i in $\mathcal{N}_1$ the optimal x_i is, from (4),

$$x_i = \frac{1}{2}(\overline{y}_1 + \overline{y}_2) + (y_i - \overline{y}_1) = y_i - \frac{1}{2}(\overline{y}_1 - \overline{y}_2), \tag{5a}$$

and for i in the second tie-block

$$x_i = \frac{1}{2}(\overline{y}_1 + \overline{y}_2) + (y_i - \overline{y}_2) = y_i + \frac{1}{2}(\overline{y}_1 - \overline{y}_2). \tag{5b}$$

In multidimensional scaling all y_i are typically non-negative. From (5b) the x_i with i in the second tie block are then also non-negative. But, from (5a), if $\overline{y}_1 > \overline{y}_2$ and i is in the first tie-block we have $x_i < 0$ for all i with $y_i < \frac{1}{2}(\overline{y}_1 - \overline{y}_2)$. Suppose for example $y = (1, 9, 1, 3)$, with the first two indices in the first tie-block and the last two in the second tie-block. Then $\overline{y}_1 = 5$ and $\overline{y}_2 = 2$, and thus $\frac{1}{2}(\overline{y}_1 - \overline{y}_2) = 1.5$ and the monotone regression is $(-.5, 7.5, 2.5, 4.5)$.

If a monotone regression in any smacof iteration produces negative disparities, then the majorization step in the next iteration may fail, and convergence of the algorithm is no longer guaranteed. There are ways to get around this problem (Heiser (1991)), but they involve a more complicated majorization procedure and lead to slower convergence. And this more complicated majorization step has not been implemented in any of the techniques in either the smacof or the smacofx packages.

3 Discussion

The title of this paper somewhat overstates the case against the tertiary approach. There is nothing inherently wrong with it, and if there are not many ties it will work as well as the primary and secondary approach. We could have, for example, some kind of design which produces tie-blocks of size two. But with many ties per block and only a few blocks we will tend to get almost perfect fits, and they may not really be the type of fit we want. In MDS we generally want the individual dissimilarities to fit the corresponding distances, which means we surely do not want the type of Shepard plots like the one for the tertiary approach in the morse data analysis. Without the corresponding plots reporting an almost perfect fit can be very misleading.

References

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

De Leeuw, J., and P. Mair. 2009. "Multidimensional Scaling Using Majorization: SMACOF in R." *Journal of Statistical Software* 31 (3): 1–30. https://www.jstatsoft.org/article/view/v031i03.

Heiser, W. J. 1991. "A Generalized Majorization Method for Least Squares Multidimensional Scaling of Pseudodistances that May Be Negative." *Psychometrika* 56 (1): 7–27.

Kruskal, J. B. 1964a. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika* 29: 1–27.

———. 1964b. "Nonmetric Multidimensional Scaling: a Numerical Method." *Psychometrika* 29: 115–29.

Mair, P., P. J. F. Groenen, and J. De Leeuw. 2022. "More on Multidimensional Scaling in R: smacof Version 2." *Journal of Statistical Software* 102 (10): 1–47. https://www.jstatsoft.org/article/view/v102i10.

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

Rusch, Thomas, Jan de Leeuw, Lisha Chen, and Patrick Mair. 2025. *Smacofx: Flexible Multidimensional Scaling and 'Smacof' Extensions*. https://doi.org/10.32614/CRAN.package.smacofx.