Non-metric Multidimensional Scaling (NMDS)

- *Objectives:*

Discuss Steps for Analysis: Advantages / Disadvantages

Go over output and interpretation of Autopilot Analysis

NMDS / NMS – What is it ?

- Output:

Representation of relationships between objects (samples, species) and descriptors (environmental variables) in a reduced number of dimensions (axes)
Just like PCA

- Non-metric:

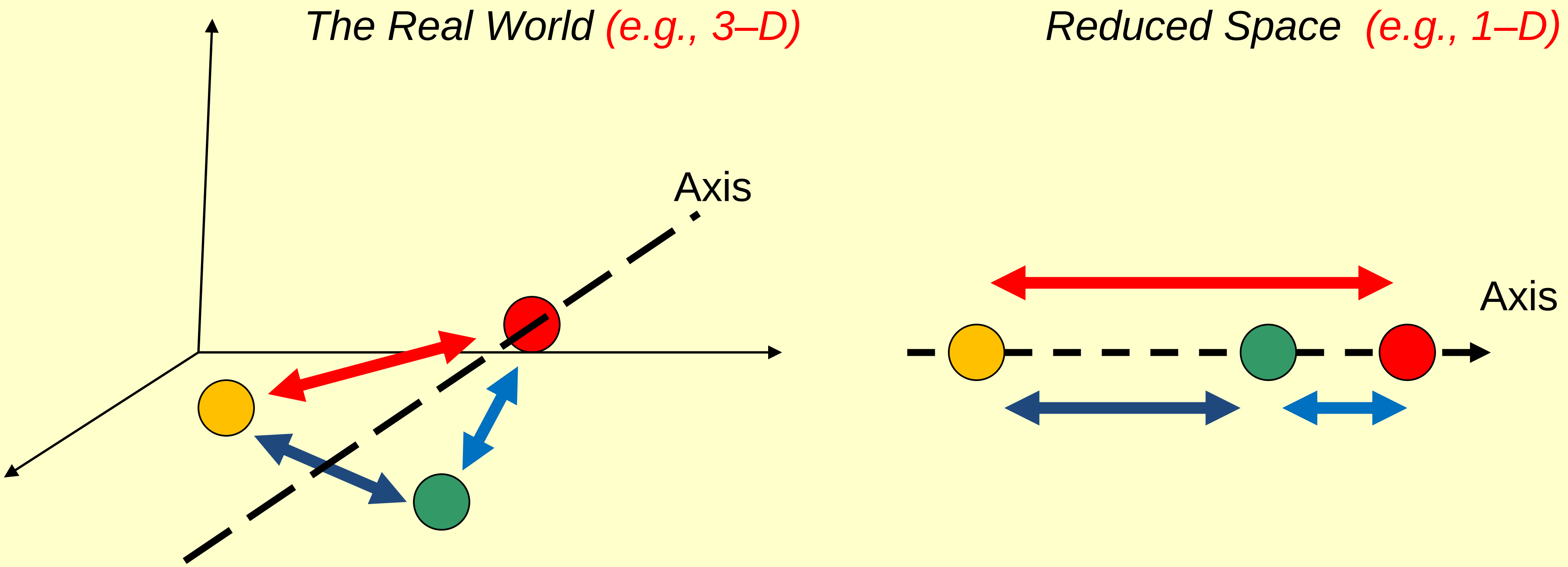
Non-parametric data analysis (ranks)
Relationships between pair-wise distances of objects (real space) and dissimilarities (ordination space) not linear

- So what ?

Axes do not correspond to eigenvectors
Unlike PCA, cannot deduce linear contribution (loadings) of various objects to the described axes

NMDS – How does it work ?

- NMDS searches for best position of n objects on k dimensions (axes) to minimize “stress” of the resulting k-dimensional configuration
- Compares the pair-wise distances (difference) of the objects in reduced ordination space (expressed in terms of axes) and the dissimilarity of the objects in the real world (expressed in terms of samples / species / variables):



The Real World (e.g., 3–D)

Reduced Space (e.g., 1–D)

NMDS – How does it work ?

- Approach:

Iterative procedure

Manipulates coordinates of pairs of observations so they fit as closely as possible measured object similarities

- Mechanics:

Using a random initialization, NMDS uses multiple iterations to find a robust pattern

Goodness of fit measured using stress, which relates pairwise distances between objects in reduced ordination space to their dissimilarities in full variable space (real world)

What does it mean to have a robust answer?

- Robustness:

Defined as:

“the persistence of a system’s characteristic behavior under perturbations or unusual or conditions of uncertainty”

- In statistics:

A robust statistical technique performs well even if its assumptions are somewhat violated by the true model from which the data were generated

So… the answer will be the same, regardless of the initial conditions (e.g., the measurements, the assumptions)

NMDS – Exploratory Method

A fun analogy: finding the lowest elevation in mars

RULES ?

- Explore around (randomly)

- Move down hill

- If you cannot go any deeper … stop

Can We Foresee Any Problems?

Local Minima

NMDS – The Good

- Being based on ranked distances, it tends to linearize the relationship between environmental / species distances (just like the Spearman Rank correlation)

- Can deal with any distance measure, data normalization and transformation

- Can handle non-metric, semi-quantitative and subjective data (e.g., best / good / bad, beaufort sea state)

- Solves the “zero truncation problem” because it does not rely on normal data

- Empirical studies have shown that:

- Use of ranks makes NMDS robust even if relationships between distances and dissimilarities are not linear

- NMDS provides appropriate summary of pair-wise distances with small number of dimensions

NMDS – The Bad

- May fail to find the global solution (minimum global stress) because of multiple local minima

- Need to account for random start of iterative process
(e.g., repeat analysis to see if random start matters)

- Computationally intensive

- Does not provide “loadings” for axes

- For a given number of dimensions, the solution for a particular axis is unique
(First dimension in 2-D solution not the same as first dimension in 3-D or 1-D)

- Axis sequence (numbers) is arbitrary
(Percent variance on a given axis does not decrease with increasing axis number)

- NMDS has difficulties in detecting discontinuities in distributions
(Remember, species abundances are ranked)

NMDS – Approach

1. Calculate dissimilarity matrix (▲) of real data.

2. Assign sample units to starting configuration in *k*-space (define initial **X**). Starting locations (scores on axes) are assigned with a random number generator.

3. Normalize **X** by subtracting axis means for each axis and dividing by overall standard deviation of scores:

$$\text{normalized } x_{il} = \frac{x_{il} - \bar{x}_l}{\sqrt{\sum_{l=1}^{k} \sum_{i=1}^{n} \left(x_{il} - \bar{x}_l\right)^2 / (n \bullet k)}}$$

(n = samples, k = dimensions)

NMDS – Approach

4. Calculate **D** using the Euclidean distances between sample units in *k*-space.

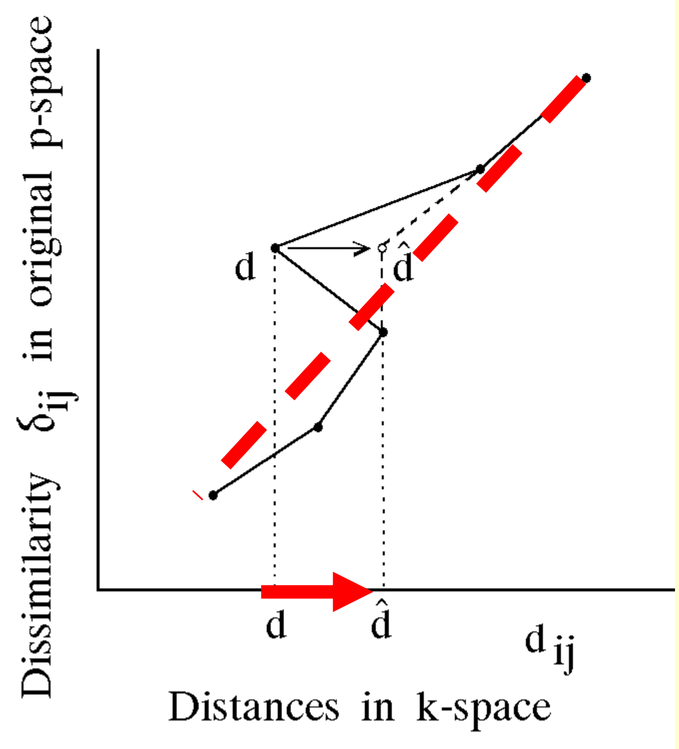
5. Rank elements of ▲ in ascending order.

6. Put the elements of **D** in the same order as ▲.

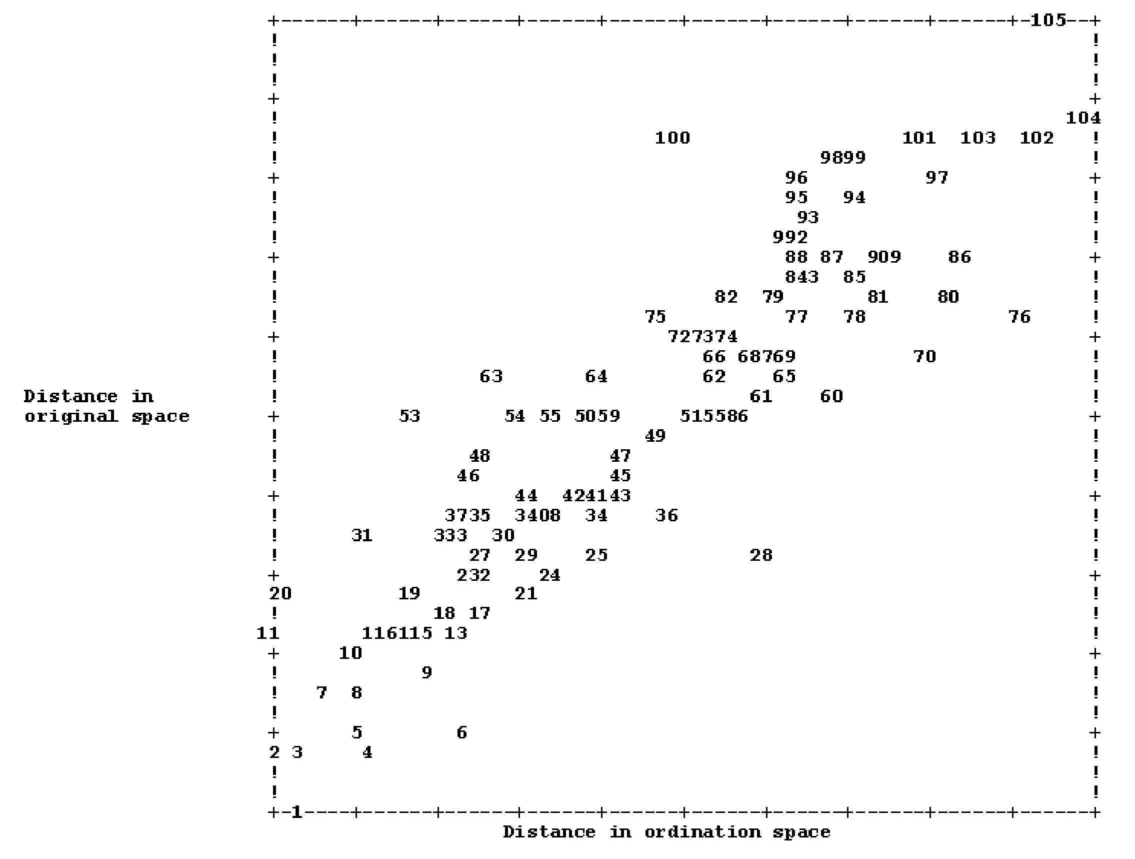
7. Calculate $\hat{d}_{ij}$
Created by replacing elements of **D** which do not meet monotonicity, with elements $\hat{\mathbf{D}}$.

Software plot sample pair-wise dissimilarities (y axis) versus distances in k-space (x axis)

Stress is based on the distances in k-space



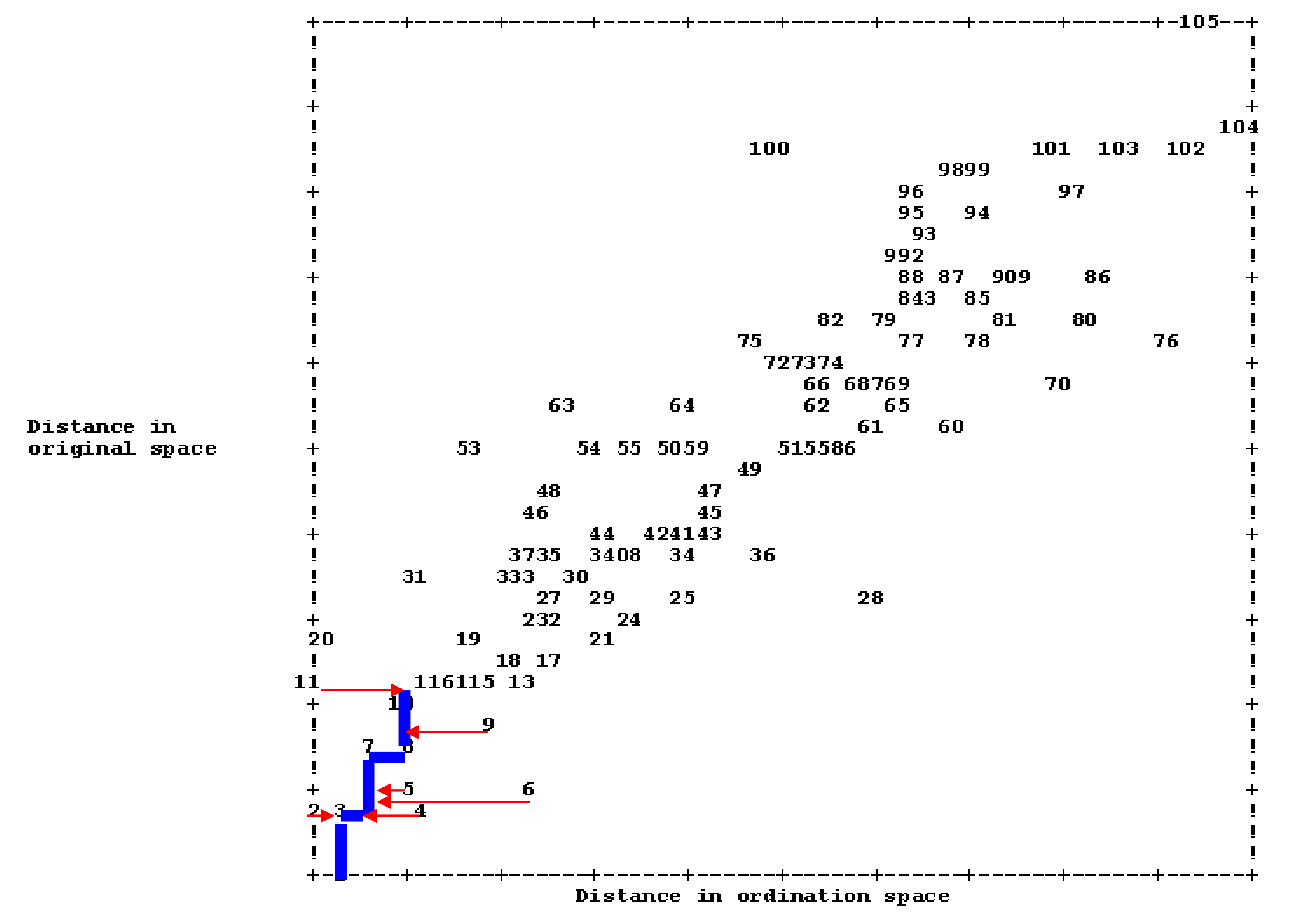
NMDS – Approach



```
                     +------+------+------+------+------+------+------+------+------+-105--+
                     !                                                                     !
                     !                                                                     !
                     !                                                                     !
                     +                                                                     +
                     !                                                                   104
                     !                                100                    101  103  102   !
                     !                                              9899                   !
                     +                                           96          97            +
                     !                                           95   94                   !
                     !                                            93                       !
                     !                                          992                        !
                     +                                           88 87  909    86          +
                     !                                           843  85                   !
                     !                                     82  79       81    80           !
                     !                               75          77   78            76     !
                     +                                 727374                              +
                     !                                    66 68769          70             !
                     !                 63       64        62    65                         !
Distance in          !                                        61    60                     !
original space       +          53       54 55 5059     515586                             +
                     !                               49                                    !
                     !                27          47                                       !
                     !               46           45                                       !
                     +                    44  424143                                       +
                     !              3735  3408  34    36                                   !
                     !      31     333  30                                                 !
                     !                27  29    25            28                           !
                     +               232    24                                             +
                     20         19        21                                               !
                     !             18 17                                                   !
                    11       116115 13                                                     !
                     +     10                                                              +
                     !            9                                                        !
                     !   7  8                                                              !
                     !                                                                     !
                     +      5        6                                                     +
                     2 3     4                                                             !
                     !                                                                     !
                     !                                                                     !
                     +-1----+------+------+------+------+------+------+------+------+------+
                                         Distance in ordination space
```

Plot of distance in ordination k-space (horizontal axis) vs. dissimilarity in original *p*-dimensional space (vertical axis). Points are labeled with the ranked distance (dissimilarity) in the original space.

NMDS – Approach



```
                     +------+------+------+------+------+------+------+------+------+-105--+
                     !                                                                     !
                     !                                                                     !
                     !                                                                     !
                     +                                                                     +
                     !                                                                   104
                     !                                100                  101  103  102   !
                     !                                              9899                   !
                     +                                           96          97            +
                     !                                           95   94                   !
                     !                                            93                       !
                     !                                          992                        !
                     +                                           88 87  909    86          +
                     !                                           843  85                   !
                     !                                     82  79       81    80           !
                     !                               75          77   78            76     !
                     +                                 727374                              +
                     !                                    66 68769          70             !
                     !                 63       64        62    65                         !
Distance in          !                                        61    60                     !
original space       +          53       54 55 5059     515586                             +
                     !                               49                                    !
                     !                48          47                                       !
                     !               46           45                                       !
                     +                    44  424143                                       +
                     !              3735  3408  34    36                                   !
                     !      31     333  30                                                 !
                     !                27  29    25            28                           !
                     +               232    24                                             +
                     20         19        21                                               !
                     !             18 17                                                   !
                    11       116115 13                                                     !
                     +     10                                                              +
                     !            9                                                        !
                     !   7  8                                                              !
                     !                                                                     !
                     +      5        6                                                     +
                     2 3     4                                                             !
                     !                                                                     !
                     !                                                                     !
                     +------+------+------+------+------+------+------+------+------+------+
                                         Distance in ordination space
```

8. Calculate d terms: shifts in k-dimensional distances (x axis) to reach monotonic (gradual) change in distances in the original data (y axis)

NMDS – Stress

9. Calculate raw stress, S*

$$S^* = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (d_{ij} - \hat{d}_{ij})^2$$

Note: *S** measures the departure from monotonicity.

If *S** = 0, the relationship is perfectly monotonic.

NMDS – Stress

10. Because raw stress is altered if the configuration of points changes (e.g., point locations, number dimensions) it is necessary to standardize ("normalize") stress.

Kruskal's stress formula one:

$$S = S^* / \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij}^{\ 2}$$

PC-ORD reports S_R, the square root of scaled stress:

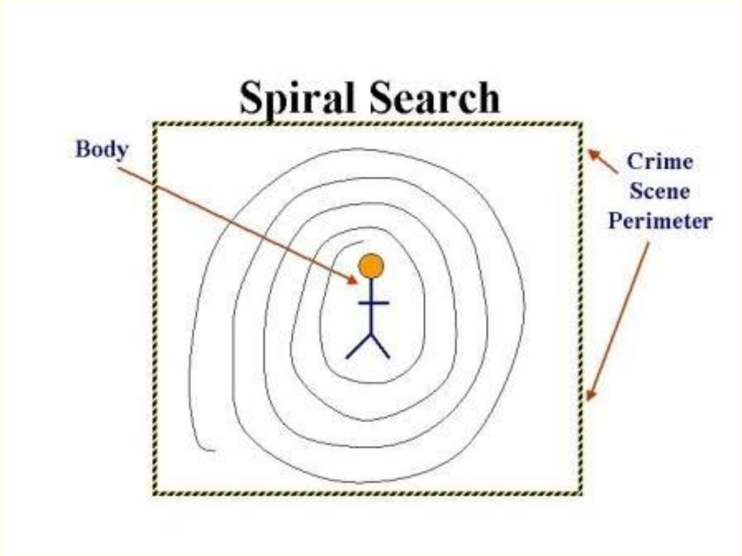
Analogous to the standard deviation of stress.

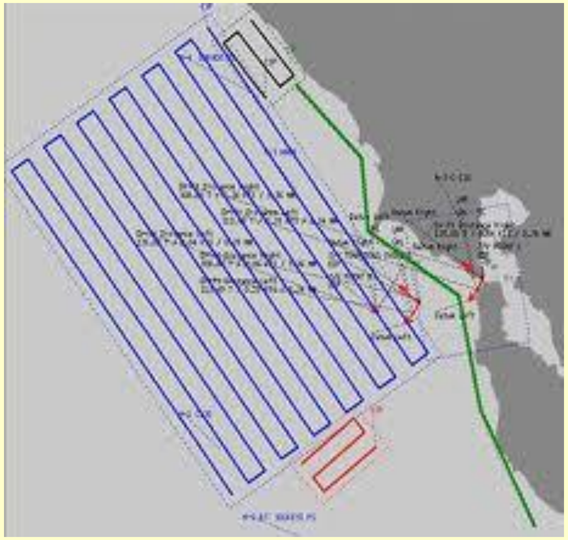
Multiplied by 100 to rescale the result from 0 to 100:

$$S_R = 100\sqrt{S}$$

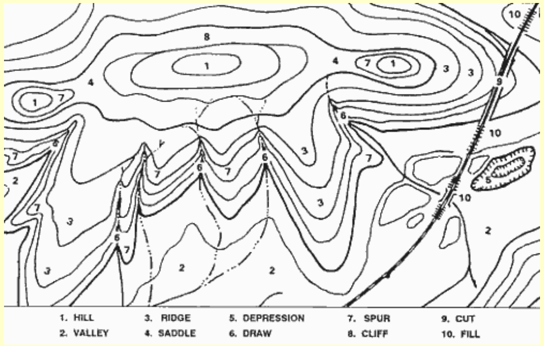
Finding the Stress Minimum

- ➢ How should the rover search the ordination landscape ?



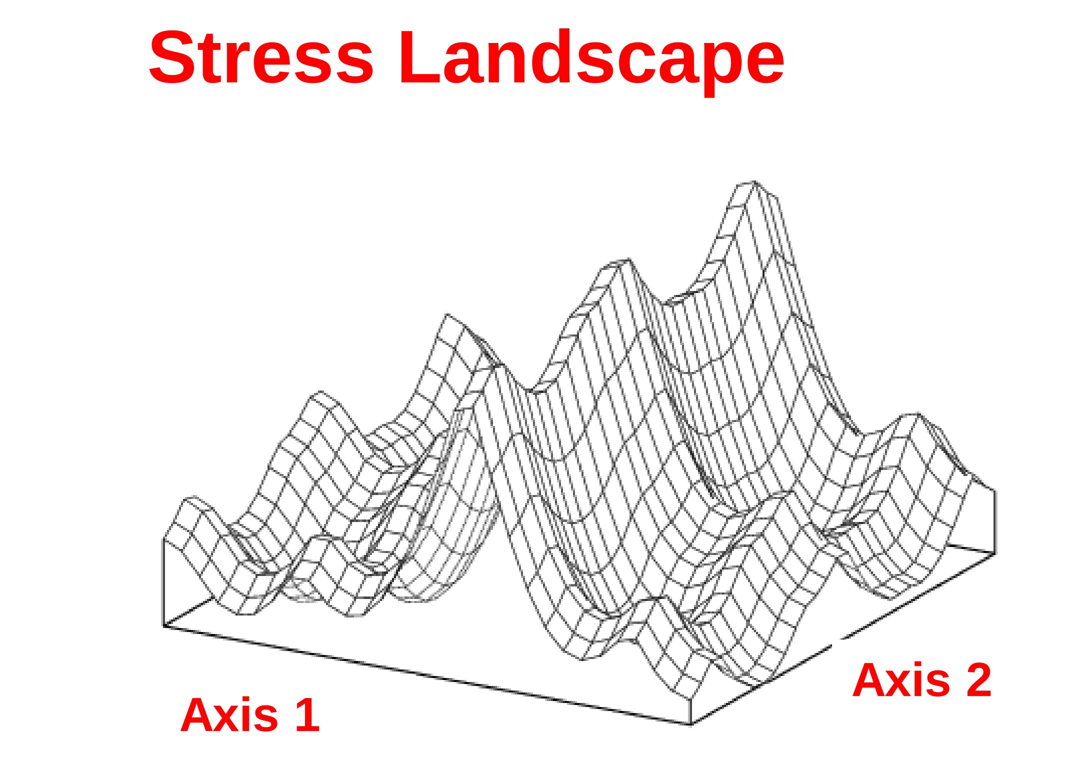


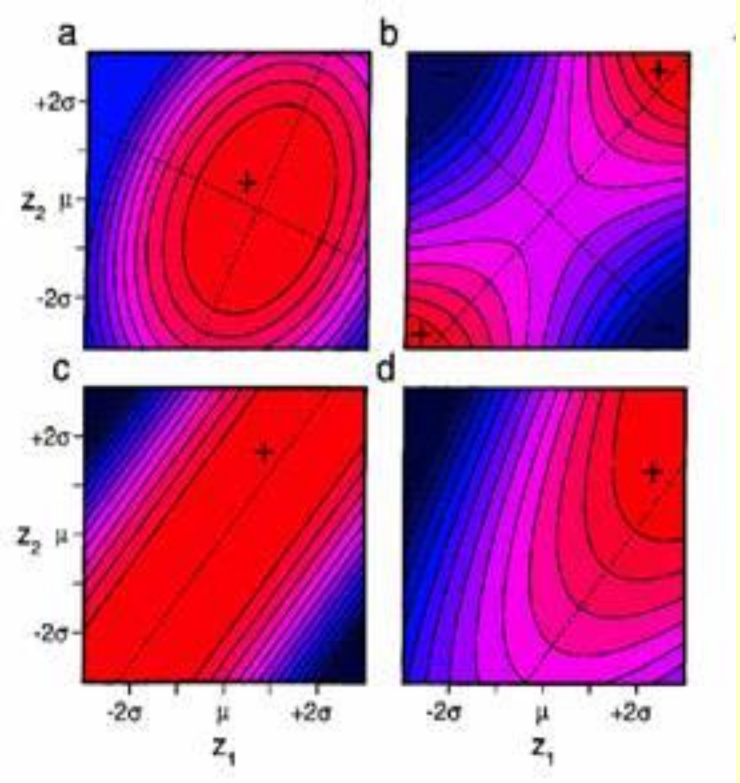
- ➢ What if “rover” read the landscape as it goes along ?



NMDS – Approach

- Crawling through the landscape in search of the optimum





The goal is to minimize stress
(to end up in a valley)

Some landscapes are trickier
than others

NMDS – Approach

11. Now the program tries to minimize *S* by changing the configuration of the sample units in the *k*-space.

Calculates negative gradient of stress for each point *i*

The amount of movement in direction of the negative gradient is set by **step length**, **a**, which is set at 0.2 initially.

The step size is recalculated after each step so it gets smaller as reductions in stress become progressively smaller.

12. Iterate (go to step 3) until either:

a set **maximum number of iterations** is reached

OR

a **criterion of stability** is met

NMS – Approach

- The starting configuration can influence the result
  - Beware of local minima (pits)
  - Avoid unstable solutions (saddle points)

- The starting configuration can be selected in two ways:
  - Use a **random** starting configuration
  - Use **coordinates** from another ordination method

Recommendation: Use a random start

- A high number of random starting configurations often provides a solution with lower stress

- This approach avoids having to decide on what other method to use – lose the great benefits of NMDS

NMDS – Approach

- Evaluate whether NMDS is extracting stronger axes than expected by chance

- Statistical Significance Based on Randomization Test (Monte Carlo approach):

p = (1+n) / (1+N) **(one tailed test)**

***n* =** number of randomized runs with final stress
less than or equal to the observed minimum stress

N = number of randomized runs

Recommendation: Use a large number of runs

- Note: Number real runs and randomized runs do not need to be equal
- We want large number of real data runs to obtain robust (minimum) answer
- We need large enough number of randomized runs to calculate the p value with the desired resolution (1000 runs for 0.001 alpha level)
- However, time intensive computational methods can take a long time

NMDS – Approach

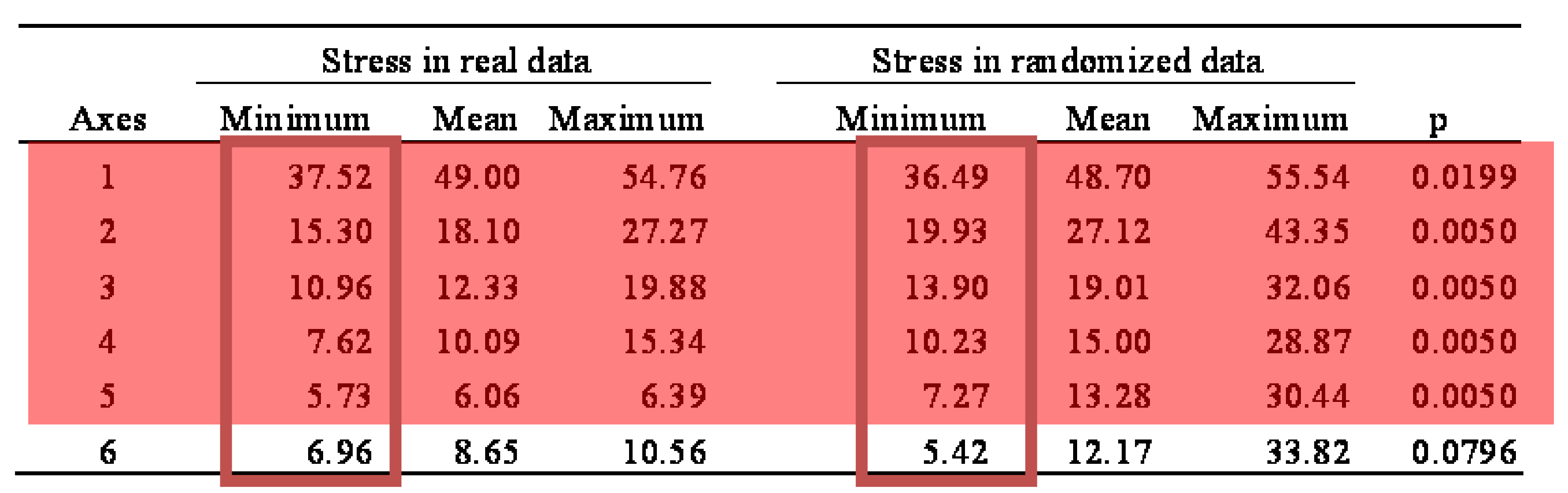
- Statistical Significance Based on Randomization Test

(p value: *p* = (1+*n*) / (1+*N*))

Axis 1: p = 0.005 = 4 / 201
Axis 2 - 5: p = 0.005 = 1 / 201
Axis 6: p = 0.0796 = 16 / 201

(50 runs)

(200 runs)



	Stress in real data			Stress in randomized data			
Axes	Minimum	Mean	Maximum	Minimum	Mean	Maximum	p
1	37.52	49.00	54.76	36.49	48.70	55.54	0.0199
2	15.30	18.10	27.27	19.93	27.12	43.35	0.0050
3	10.96	12.33	19.88	13.90	19.01	32.06	0.0050
4	7.62	10.09	15.34	10.23	15.00	28.87	0.0050
5	5.73	6.06	6.39	7.27	13.28	30.44	0.0050
6	6.96	8.65	10.56	5.42	12.17	33.82	0.0796

- Results:

Stress declines with increasing dimensions

On average, real data yield lower stress than randomized data

NMDS – Approach

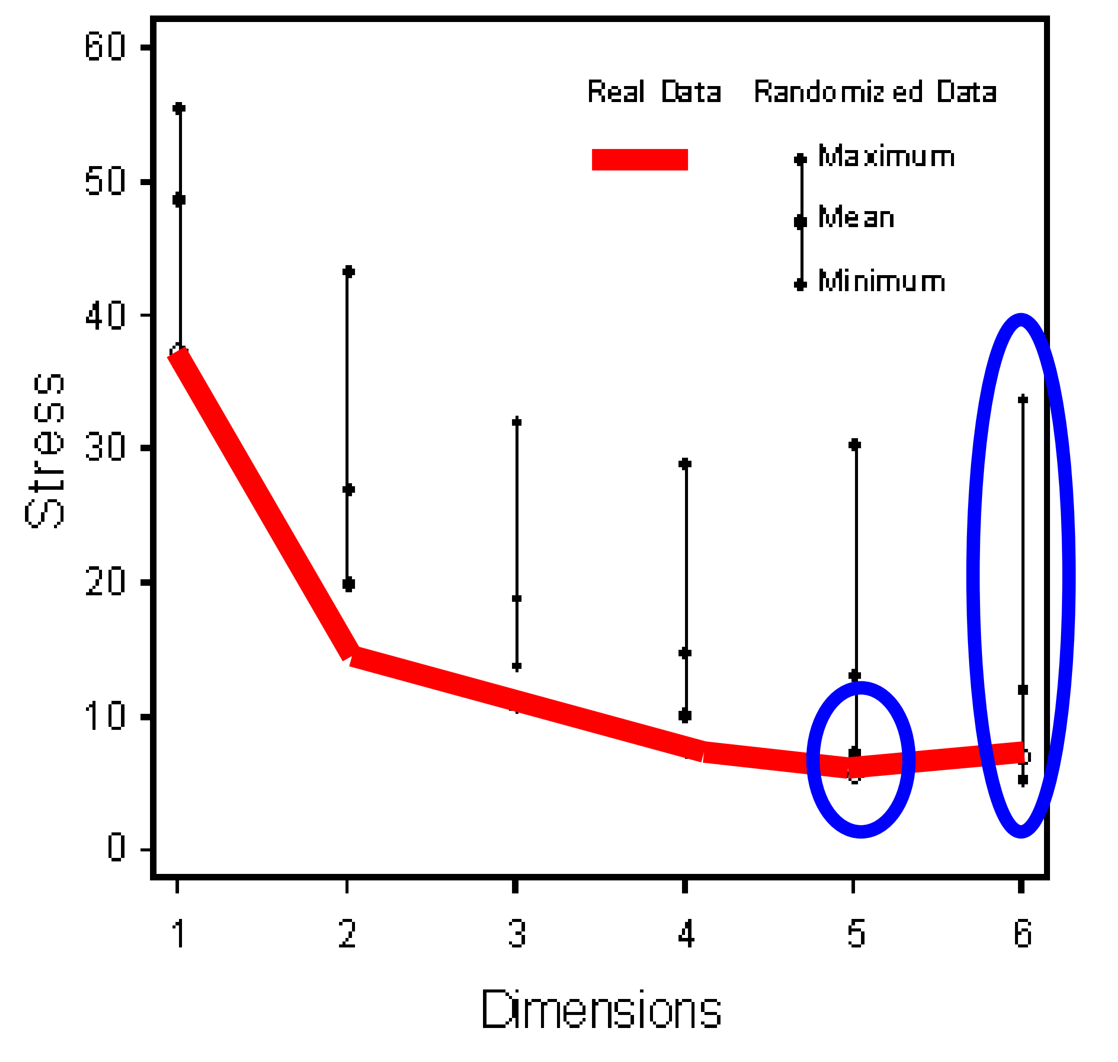
- Stress Interpretation:

- “Real Data”:

Declines with increasing dimensions (from 1 to 5)

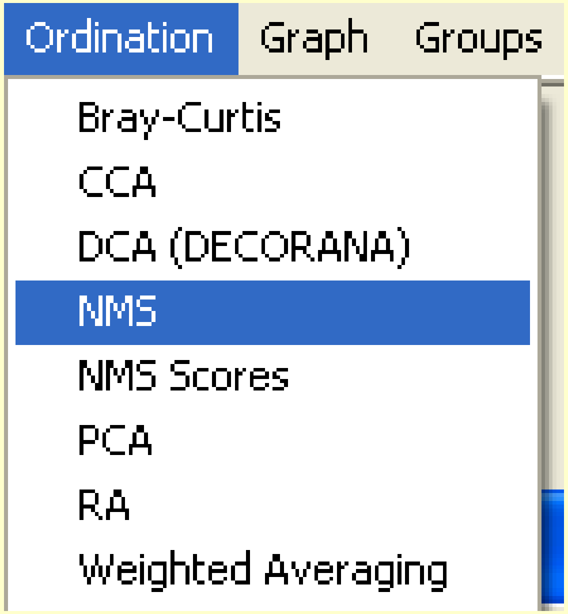
- “Randomized Data”:

Real data stress below the distribution of stress for the randomized data (for dimensions 1 to 5)



NMS – Autopilot Mode

The automatic procedure determines most appropriate dimensionality, assigns statistical significance with randomizations, and avoids local minima (using random iterations)



- Advantages:

  Uses default settings and decides number of axes for you

- Disadvantages:

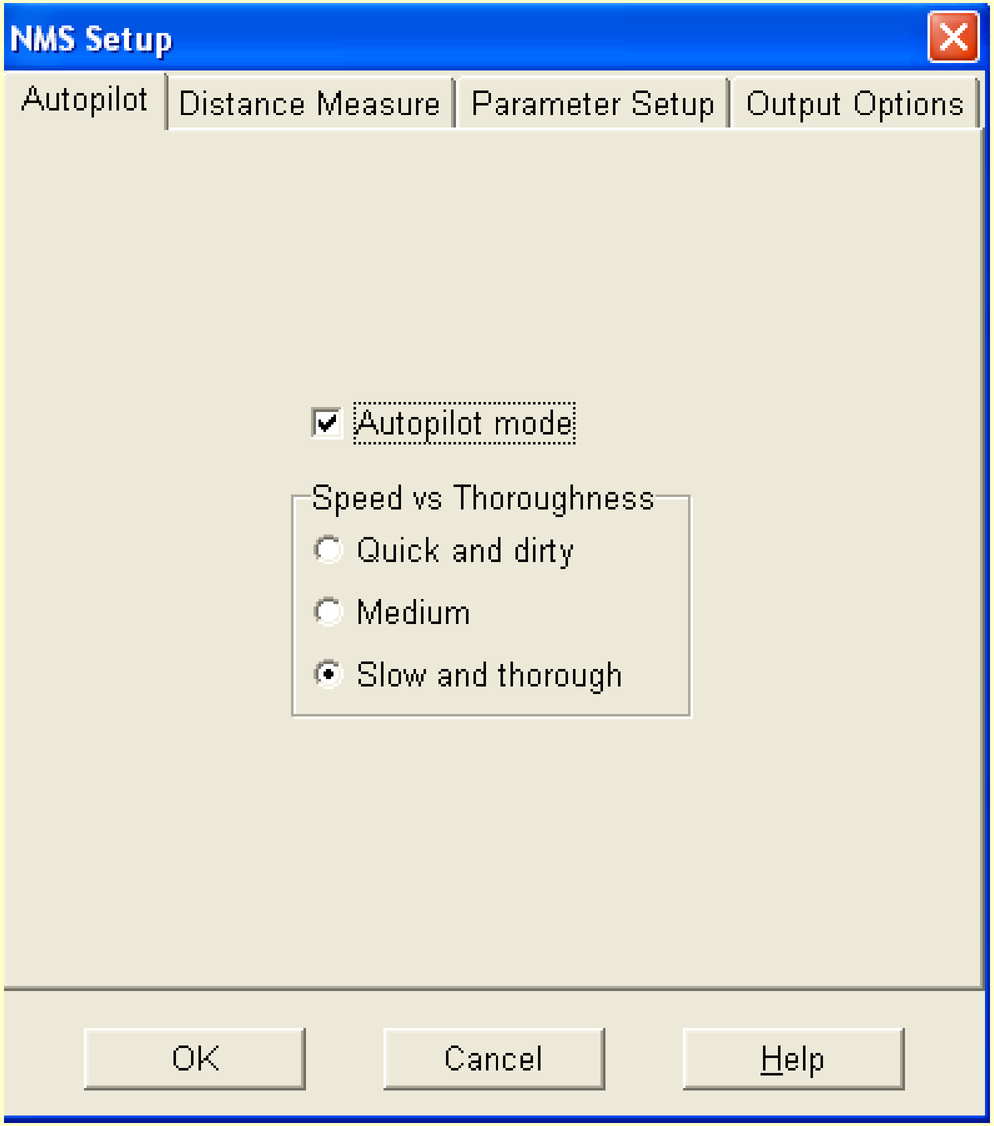
  User may want additional output products.
  User decides number of axes based on additional considerations

NMS – Autopilot Mode

- Autopilot NMS mode has three settings:

Balance:
Speed vs Thoroughness

Quick and Dirty
Medium
Slow and Thorough

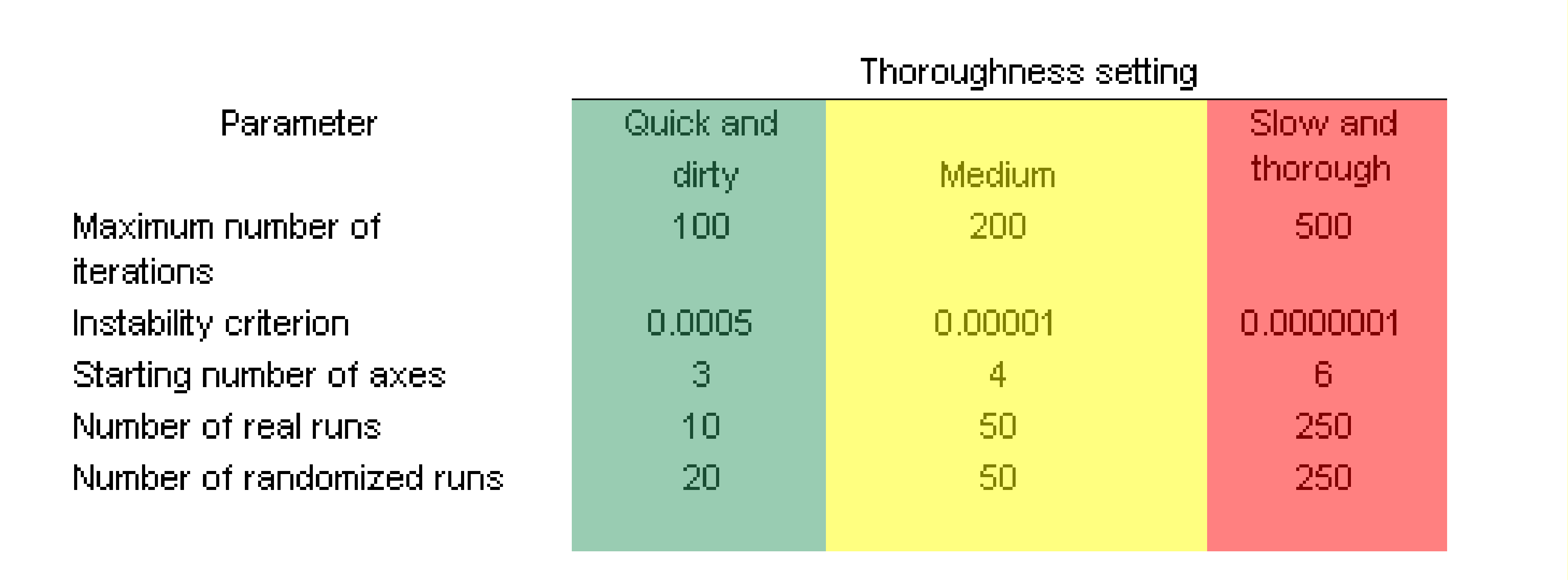


☑ Autopilot mode

Speed vs Thoroughness
○ Quick and dirty
○ Medium
● Slow and thorough

NMS – Autopilot Mode

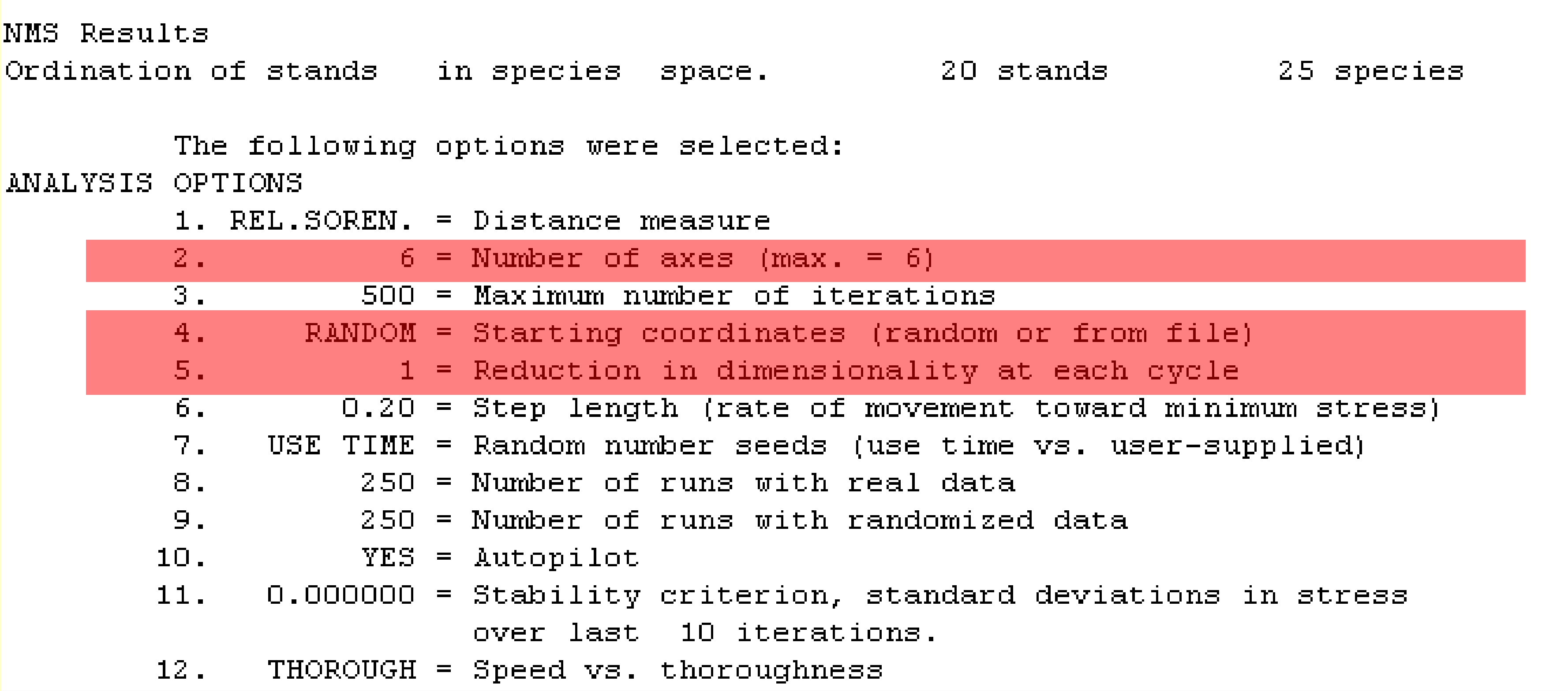
- The autopilot NMS mode provides three settings



	Thoroughness setting		
Parameter	Quick and dirty	Medium	Slow and thorough
Maximum number of iterations	100	200	500
Instability criterion	0.0005	0.00001	0.0000001
Starting number of axes	3	4	6
Number of real runs	10	50	250
Number of randomized runs	20	50	250

NMS – Results

- Examine Results.txt file: Settings / Options

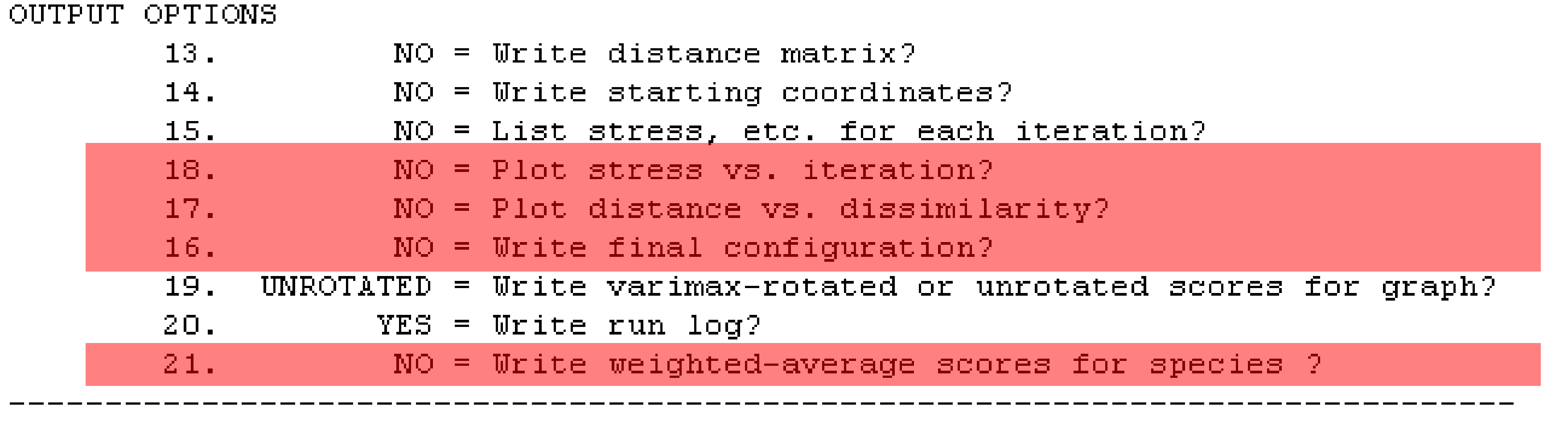


```
NMS Results
Ordination of stands   in species  space.        20 stands          25 species

         The following options were selected:
ANALYSIS OPTIONS
         1. REL.SOREN. = Distance measure
         2.          6 = Number of axes (max. = 6)
         3.        500 = Maximum number of iterations
         4.     RANDOM = Starting coordinates (random or from file)
         5.          1 = Reduction in dimensionality at each cycle
         6.       0.20 = Step length (rate of movement toward minimum stress)
         7.   USE TIME = Random number seeds (use time vs. user-supplied)
         8.        250 = Number of runs with real data
         9.        250 = Number of runs with randomized data
        10.        YES = Autopilot
        11.   0.000000 = Stability criterion, standard deviations in stress
                         over last  10 iterations.
        12.   THOROUGH = Speed vs. thoroughness
```

- Up to 6 dimensions (for sake of interpretation)
- Random start (to avoid local minima)
- Reduction in dimensionality (D: 6,5,4,3,2,1)

NMS – Results

- Examine Results.txt file: Settings / Options (all Dimensions)



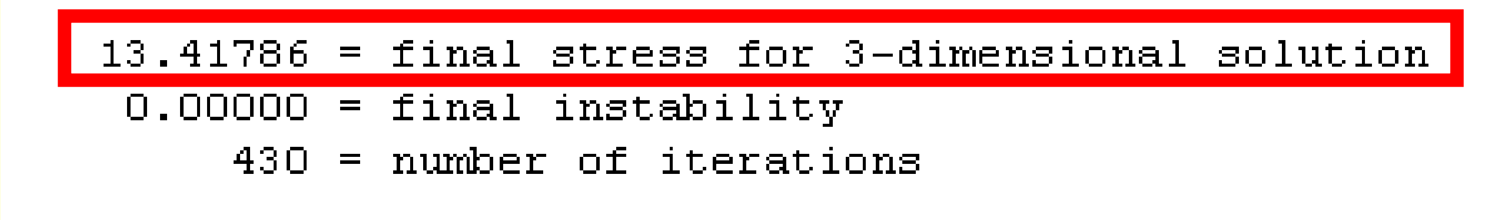
```
OUTPUT OPTIONS
        13.          NO = Write distance matrix?
        14.          NO = Write starting coordinates?
        15.          NO = List stress, etc. for each iteration?
        18.          NO = Plot stress vs. iteration?
        17.          NO = Plot distance vs. dissimilarity?
        16.          NO = Write final configuration?
        19.   UNROTATED = Write varimax-rotated or unrotated scores for graph?
        20.         YES = Write run log?
        21.          NO = Write weighted-average scores for species ?
------------------------------------------------------------------------------
```

- Cannot monitor changing stress
- Cannot assess linearity of distances / dissimilarities
- Cannot see scores for all the runs – just for final run
- Cannot see scores for species – just for final run

NMS – Results

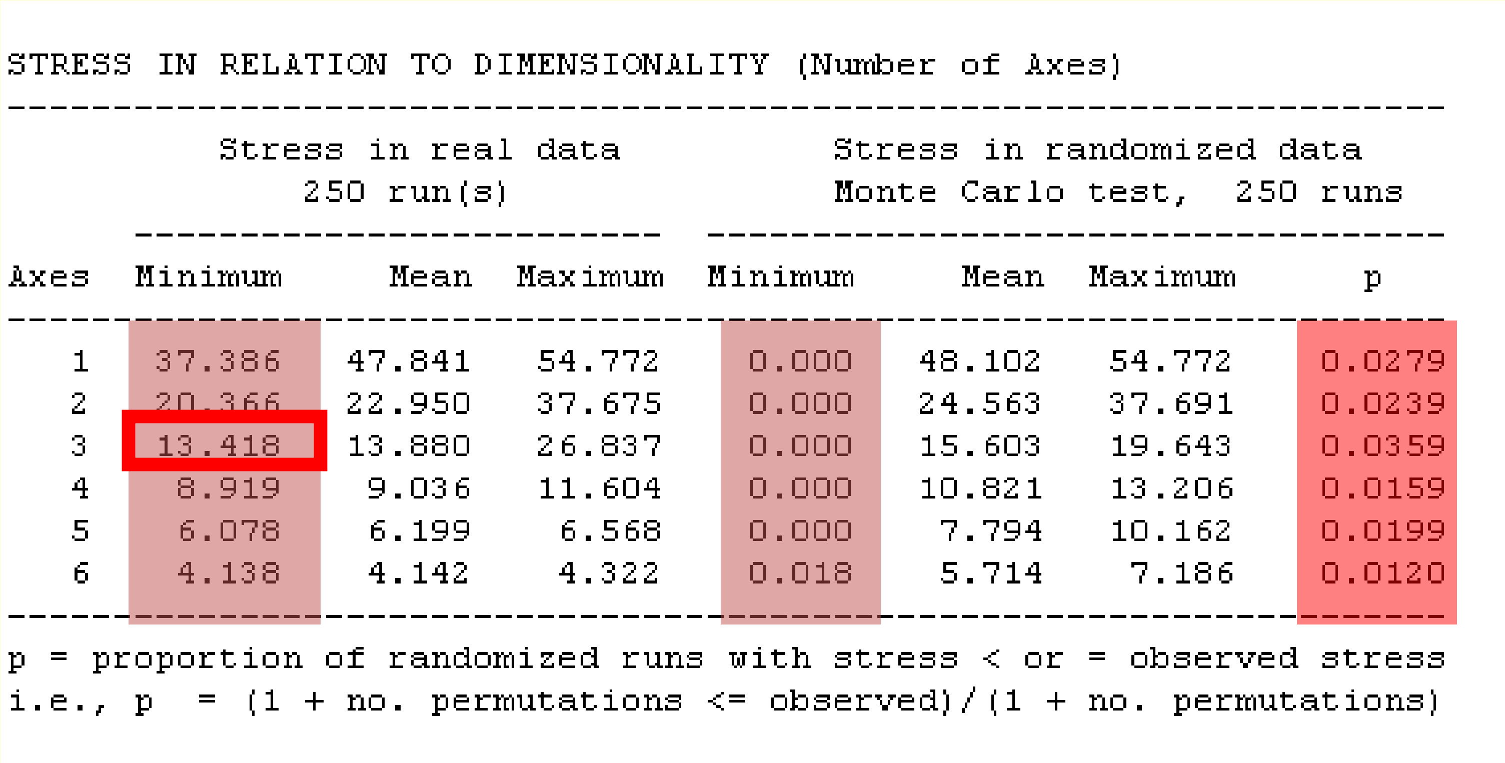
- Examine Results.txt file: Results for best result

Stress



```
 13.41786 = final stress for 3-dimensional solution
  0.00000 = final instability
      430 = number of iterations
```

P values

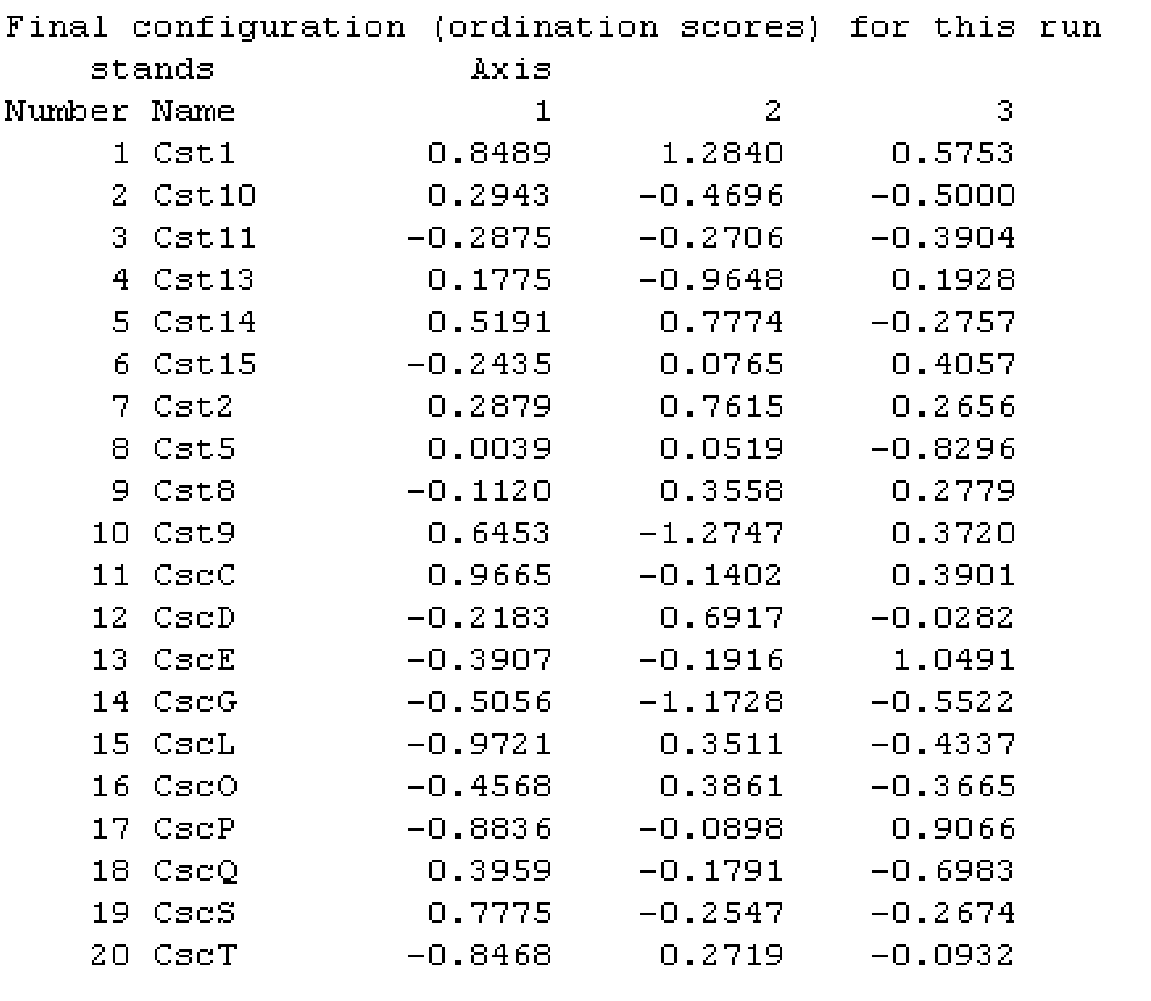


```
STRESS IN RELATION TO DIMENSIONALITY (Number of Axes)
---------------------------------------------------------------------
          Stress in real data            Stress in randomized data
              250 run(s)                 Monte Carlo test,  250 runs
      -------------------------  ----------------------------------
Axes  Minimum     Mean  Maximum  Minimum     Mean  Maximum         p
---------------------------------------------------------------------
   1   37.386   47.841   54.772    0.000   48.102   54.772    0.0279
   2   20.366   22.950   37.675    0.000   24.563   37.691    0.0239
   3   13.418   13.880   26.837    0.000   15.603   19.643    0.0359
   4    8.919    9.036   11.604    0.000   10.821   13.206    0.0159
   5    6.078    6.199    6.568    0.000    7.794   10.162    0.0199
   6    4.138    4.142    4.322    0.018    5.714    7.186    0.0120
---------------------------------------------------------------------
p = proportion of randomized runs with stress < or = observed stress
i.e., p  = (1 + no. permutations <= observed)/(1 + no. permutations)
```

NMS – Results

- Examine Results.txt file: Results for best result

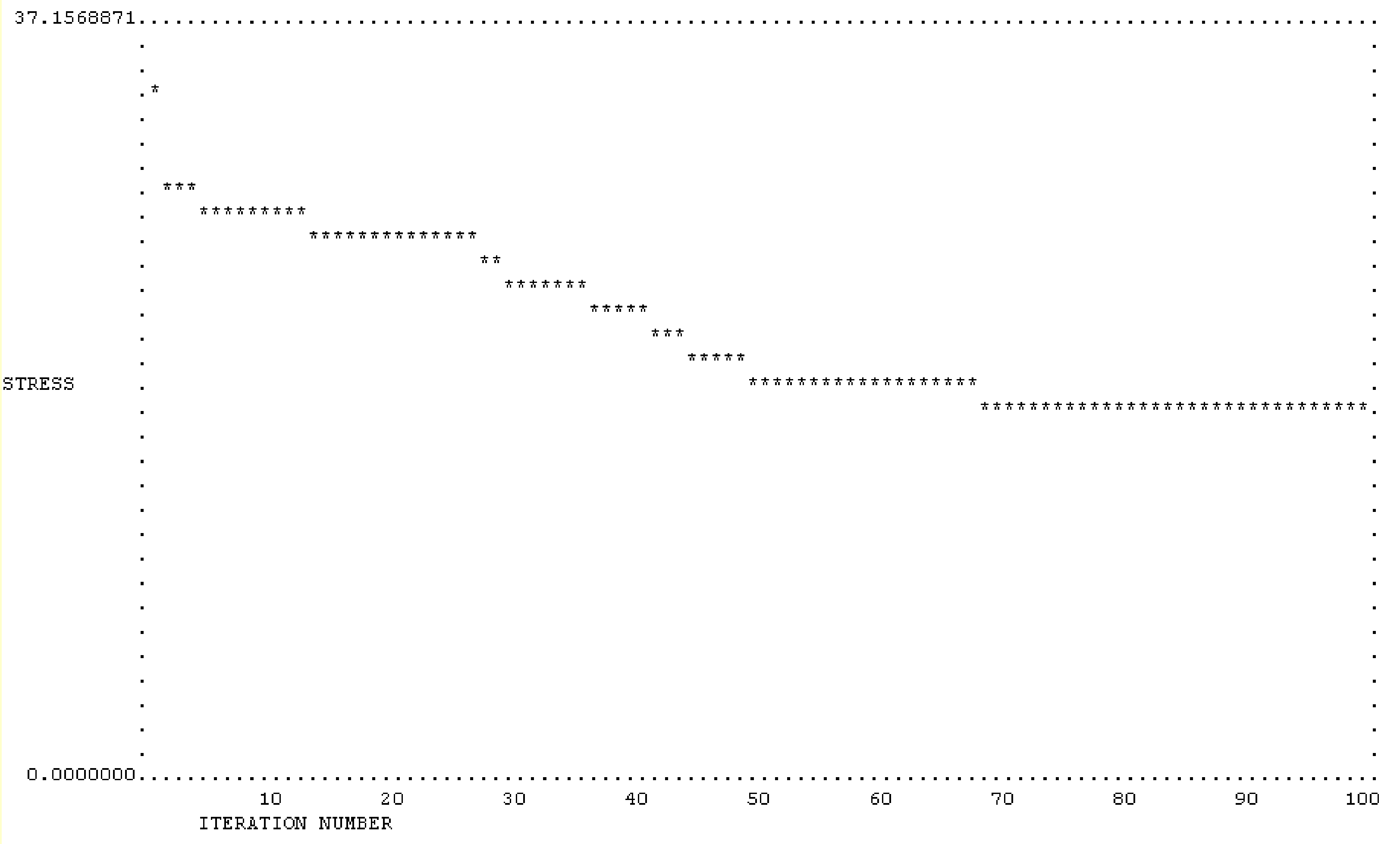
Scores



```
Final configuration (ordination scores) for this run
    stands                Axis
Number Name                  1           2           3
     1 Cst1             0.8489      1.2840      0.5753
     2 Cst10            0.2943     -0.4696     -0.5000
     3 Cst11           -0.2875     -0.2706     -0.3904
     4 Cst13            0.1775     -0.9648      0.1928
     5 Cst14            0.5191      0.7774     -0.2757
     6 Cst15           -0.2435      0.0765      0.4057
     7 Cst2             0.2879      0.7615      0.2656
     8 Cst5             0.0039      0.0519     -0.8296
     9 Cst8            -0.1120      0.3558      0.2779
    10 Cst9             0.6453     -1.2747      0.3720
    11 CscC             0.9665     -0.1402      0.3901
    12 CscD            -0.2183      0.6917     -0.0282
    13 CscE            -0.3907     -0.1916      1.0491
    14 CscG            -0.5056     -1.1728     -0.5522
    15 CscL            -0.9721      0.3511     -0.4337
    16 CscO            -0.4568      0.3861     -0.3665
    17 CscP            -0.8836     -0.0898      0.9066
    18 CscQ             0.3959     -0.1791     -0.6983
    19 CscS             0.7775     -0.2547     -0.2674
    20 CscT            -0.8468      0.2719     -0.0932
```

NMS – Results

- Examine Results.txt file: Plotting Stress vs Iteration

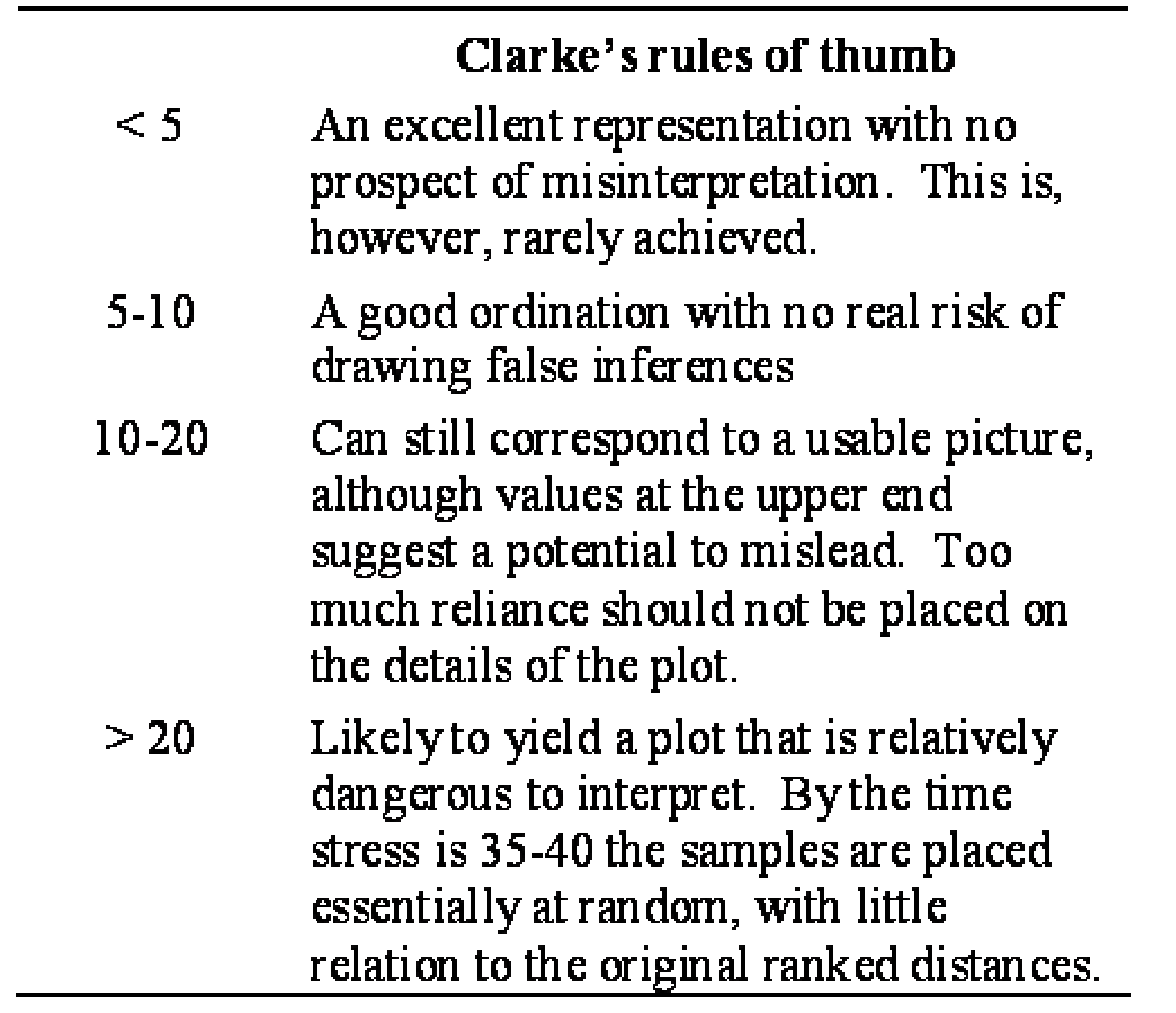


```
 37.1568871......................................................................................................
           .                                                                                                    .
           .                                                                                                    .
           .*                                                                                                   .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           . ***                                                                                                .
           .    *********                                                                                       .
           .             **************                                                                         .
           .                           **                                                                       .
           .                             *******                                                                .
           .                                    *****                                                           .
           .                                         ***                                                        .
           .                                            *****                                                   .
STRESS     .                                                 *******************                                .
           .                                                                    ********************************.
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
           .                                                                                                    .
  0.0000000......................................................................................................
                    10        20        30        40        50        60        70        80        90       100
                ITERATION NUMBER
```

- **Note:** This graph provided only for best answer (3-D)

NMS – Results

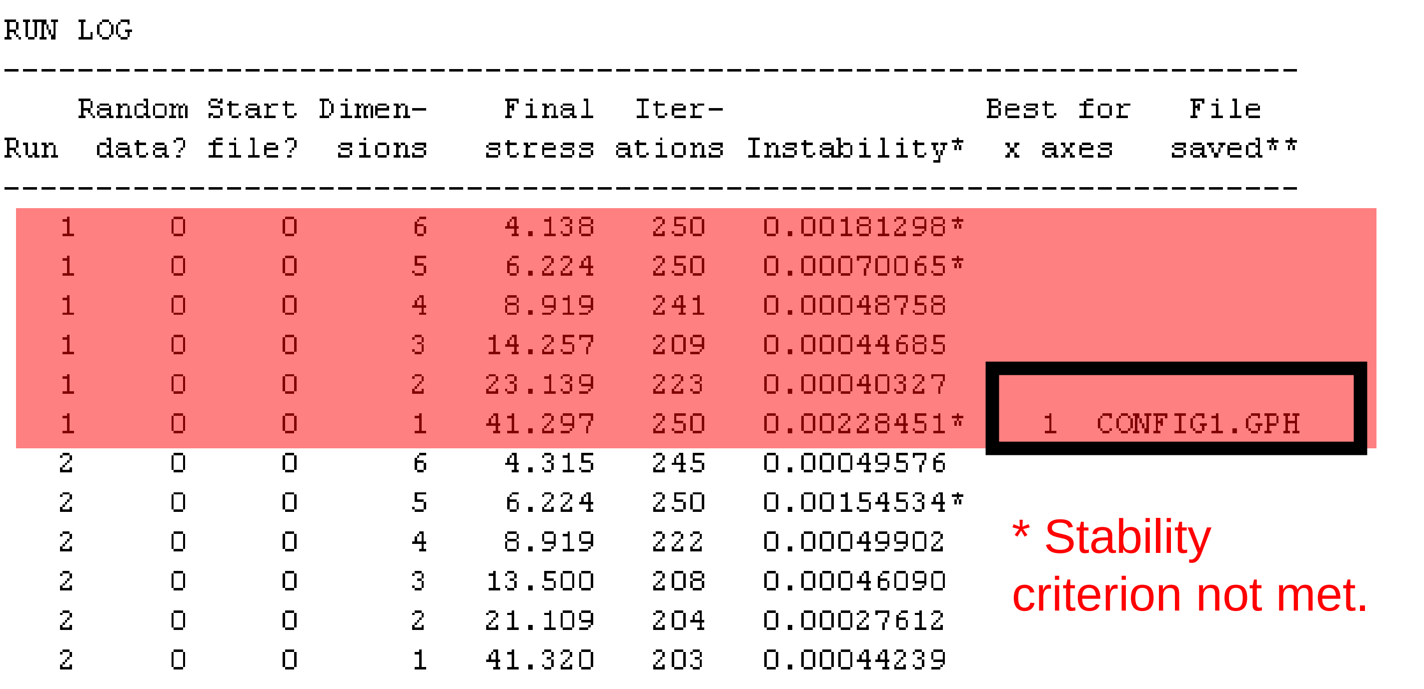
- Examine Results.txt file: Interpret Stress **(Clarke 1993)**



	Clarke's rules of thumb
< 5	An excellent representation with no prospect of misinterpretation. This is, however, rarely achieved.
5-10	A good ordination with no real risk of drawing false inferences
10-20	Can still correspond to a usable picture, although values at the upper end suggest a potential to mislead. Too much reliance should not be placed on the details of the plot.
> 20	Likely to yield a plot that is relatively dangerous to interpret. By the time stress is 35-40 the samples are placed essentially at random, with little relation to the original ranked distances.

NMS – Results

- Examine Results.txt file: Run Log



```
RUN LOG
-----------------------------------------------------------------------
    Random Start Dimen-     Final  Iter-              Best for    File
Run  data? file?  sions    stress ations Instability*  x axes    saved**
-----------------------------------------------------------------------
  1      0     0      6     4.138   250   0.00181298*
  1      0     0      5     6.224   250   0.00070065*
  1      0     0      4     8.919   241   0.00048758
  1      0     0      3    14.257   209   0.00044685
  1      0     0      2    23.139   223   0.00040327
  1      0     0      1    41.297   250   0.00228451*    1  CONFIG1.GPH
  2      0     0      6     4.315   245   0.00049576
  2      0     0      5     6.224   250   0.00154534*
  2      0     0      4     8.919   222   0.00049902
  2      0     0      3    13.500   208   0.00046090
  2      0     0      2    21.109   204   0.00027612
  2      0     0      1    41.320   203   0.00044239
```

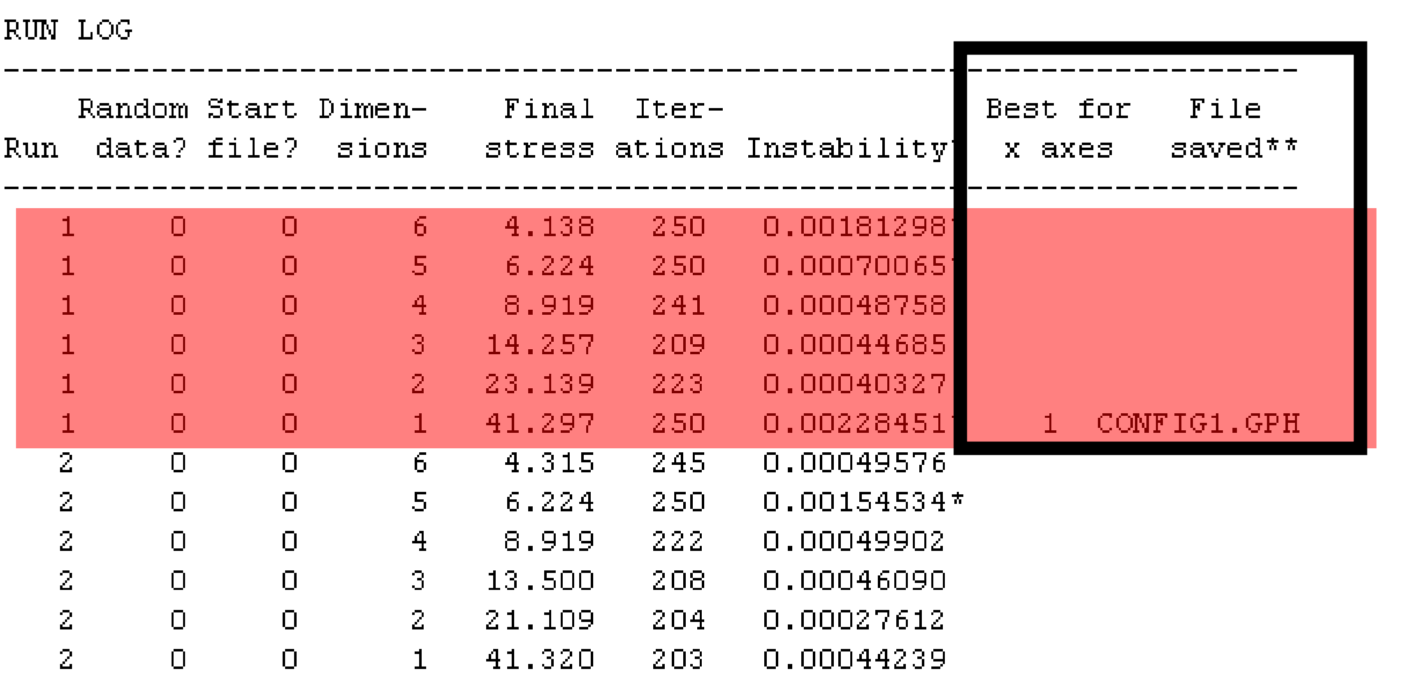
* Stability criterion not met.

Random data:
0 = not randomized,
1 = randomized

Start file:
0 = random starting coordinates
1 = read from file Seeds - initial seeds for random number generator

NMS – Results

- Examine Results.txt file: Run Log



```
RUN LOG
----------------------------------------------------------------------
    Random Start Dimen-    Final  Iter-             Best for   File
Run  data? file?  sions   stress ations Instability*  x axes   saved**
----------------------------------------------------------------------
   1     0     0      6    4.138    250   0.00181298*
   1     0     0      5    6.224    250   0.00070065*
   1     0     0      4    8.919    241   0.00048758
   1     0     0      3   14.257    209   0.00044685
   1     0     0      2   23.139    223   0.00040327
   1     0     0      1   41.297    250   0.00228451*      1  CONFIG1.GPH
   2     0     0      6    4.315    245   0.00049576
   2     0     0      5    6.224    250   0.00154534*
   2     0     0      4    8.919    222   0.00049902
   2     0     0      3   13.500    208   0.00046090
   2     0     0      2   21.109    204   0.00027612
   2     0     0      1   41.320    203   0.00044239
```

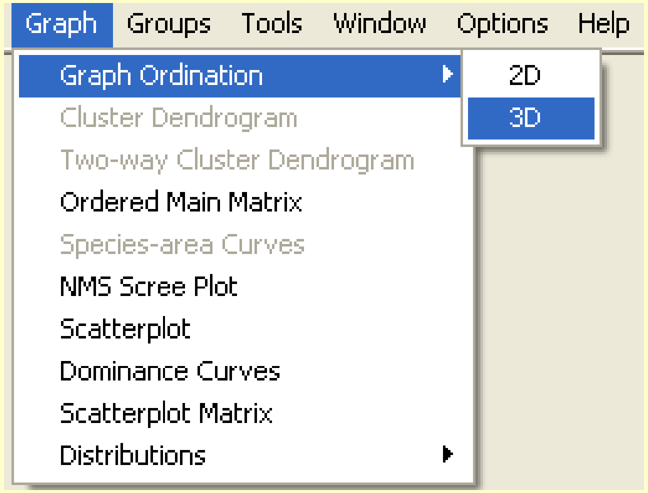
NOTE: To run single NMS ordination repeating best result,
use this file as starting configuration, rather than using random start.

Save this file with new name, to avoid overwriting it with next NMS test.

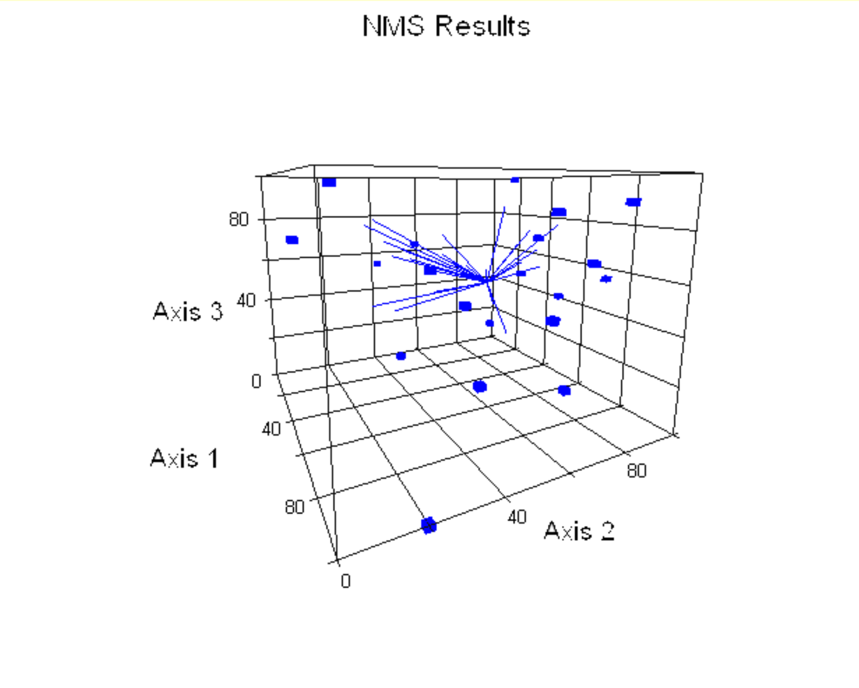
To do this: open file using File | Open | Graph Row file,
then File | Save as | Graph Row file (specify new name).

NMS – Results

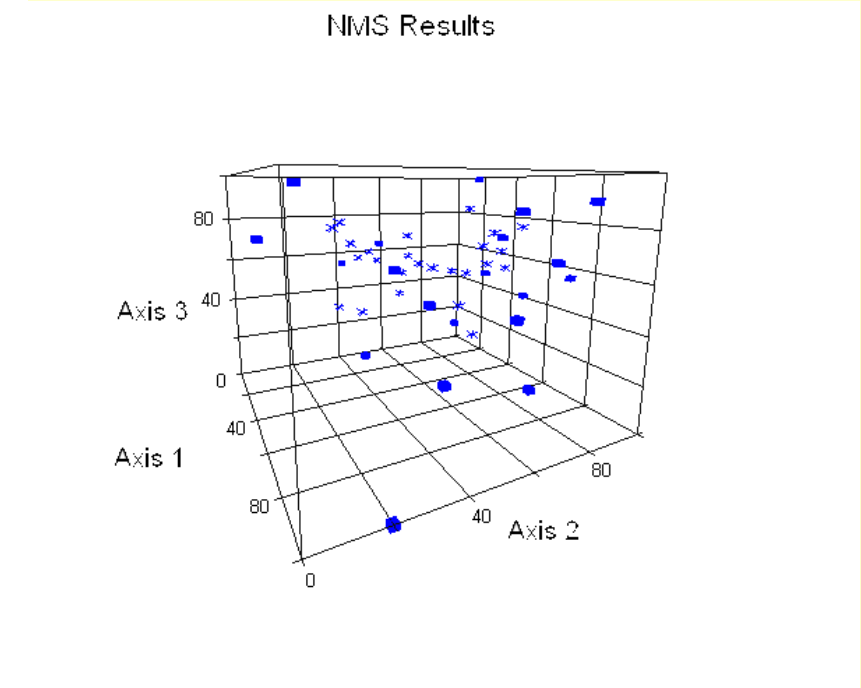
- Examine graphs: Species scores
  - Select Weighted Average Scores



Species as Vectors

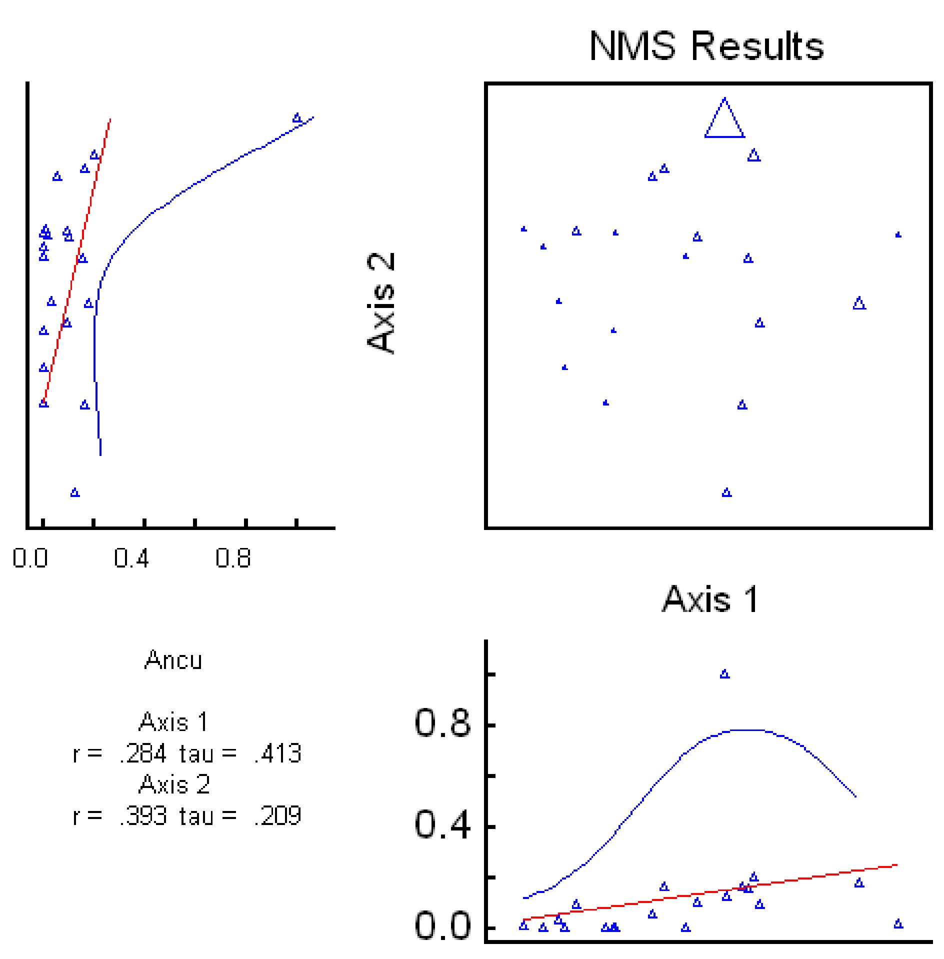


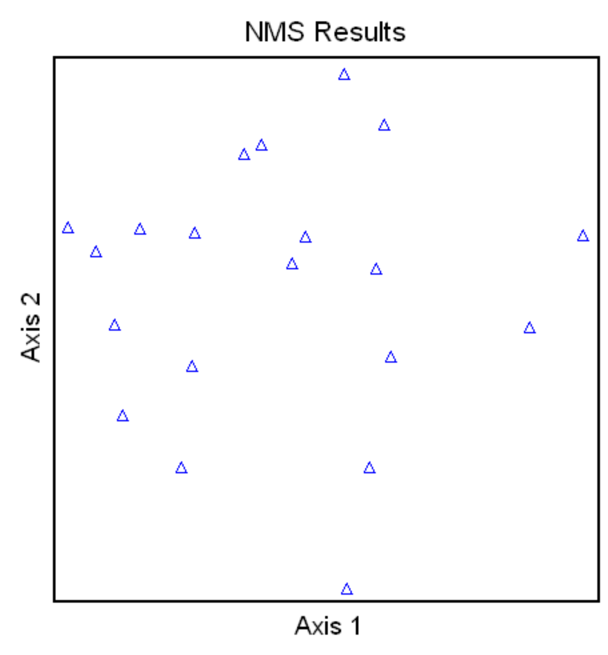
Species as Points

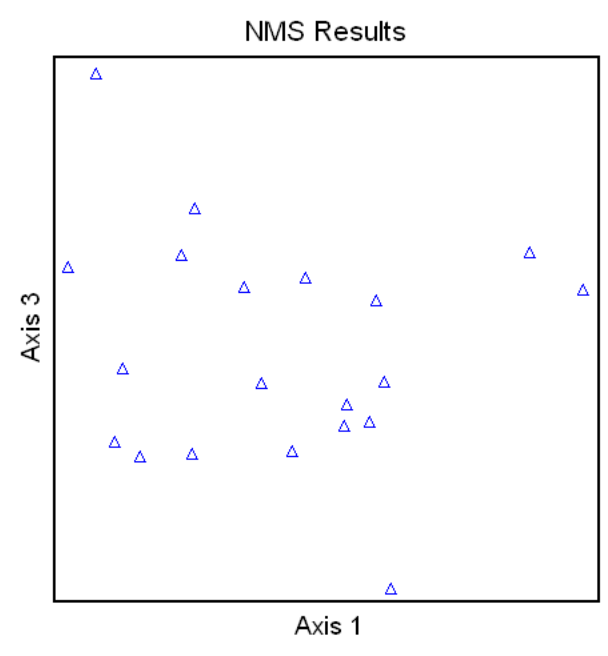


NMS – Results

- Examine graphs: 2D Ordination plots







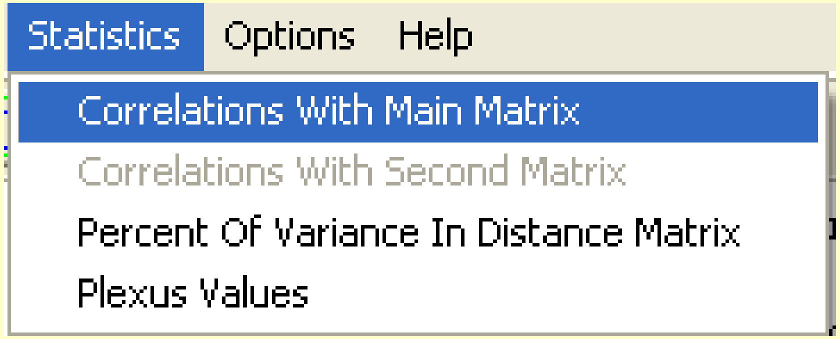
- Tau: non parametric correlation

NMS – Results

- Correlations with Matrices:

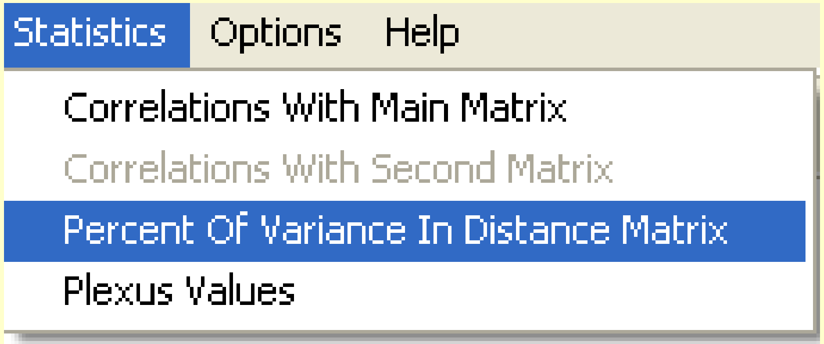
Tau (rank correlation)

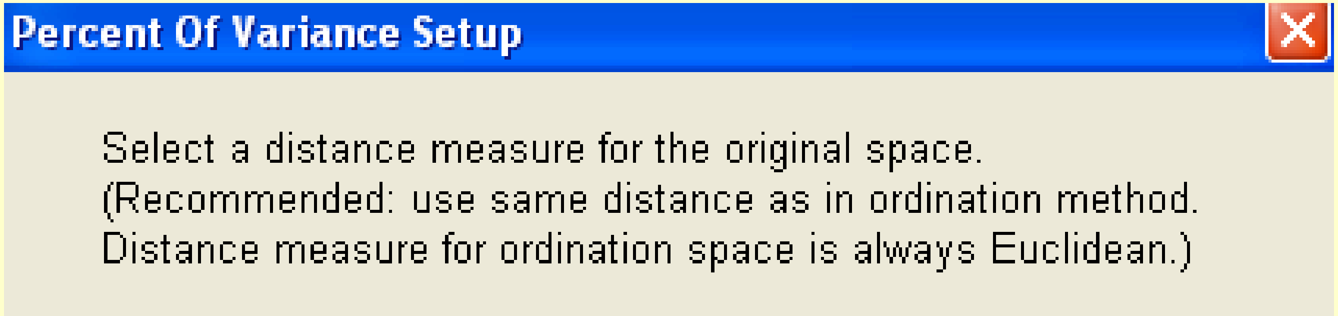
DO NOT use r ^{2} value


Statistics
Options
Help
Correlations With Main Matrix
Correlations With Second Matrix
Percent Of Variance In Distance Matrix
Plexus Values

- Percent Explained Variance:

NOTE: Use same distance metric used for NMDS analysis


Statistics
Options
Help
Correlations With Main Matrix
Correlations With Second Matrix
Percent Of Variance In Distance Matrix
Plexus Values

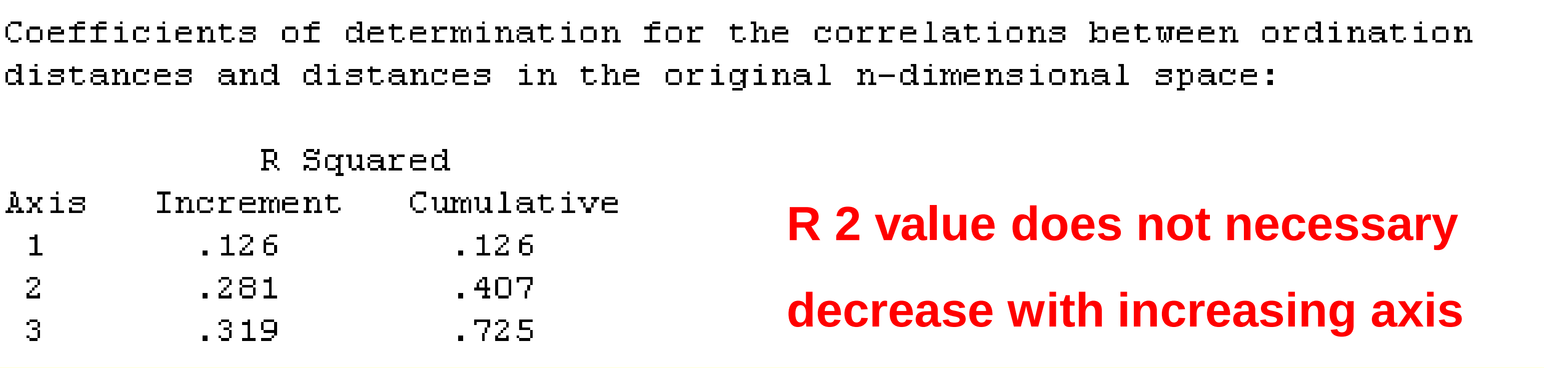


Select a distance measure for the original space.
(Recommended: use same distance as in ordination method.
Distance measure for ordination space is always Euclidean.)

NMS – Results

- Coefficient of Determination (% of Variance):

For each axis & together



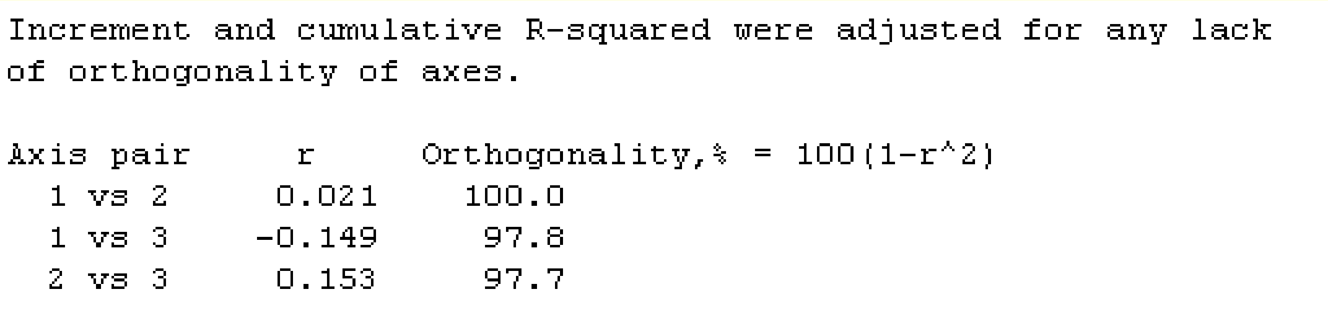
```
Coefficients of determination for the correlations between ordination
distances and distances in the original n-dimensional space:

            R Squared
Axis    Increment    Cumulative
 1         .126          .126
 2         .281          .407
 3         .319          .725
```

R 2 value does not necessary

decrease with increasing axis

- Orthogonality:

Measure independence of axes (NOTE: Not assured for NMDS)



```
Increment and cumulative R-squared were adjusted for any lack
of orthogonality of axes.

Axis pair      r     Orthogonality,% = 100(1-r^2)
  1 vs 2     0.021    100.0
  1 vs 3    -0.149     97.8
  2 vs 3     0.153     97.7
```

Project Proposal – March 13th (5 points)

1. Describe your dataset

2. Describe your “big picture” rationale for analyzing this dataset: Outline the empirical / theoretical background which stimulates the analysis. This entails using the literature provided for the course, augmented with additional relevant references. Write 1 paragraph. (+1 point for rationale and +1 point for references).
(NOTE: Use 5 – 10 references).

3. Describe your analysis approach: Outline the goal of this analysis in plain words and provide a hypothesis. Be as specific as you can, at this stage. For instance, explain whether you wish to organize your data into discrete groups, develop independent synthetic variables, merge environmental data with species records, …
Write 1 paragraph. (+1 point for hypothesis and +1 point for goal).