

# NMDS – Application & Examples

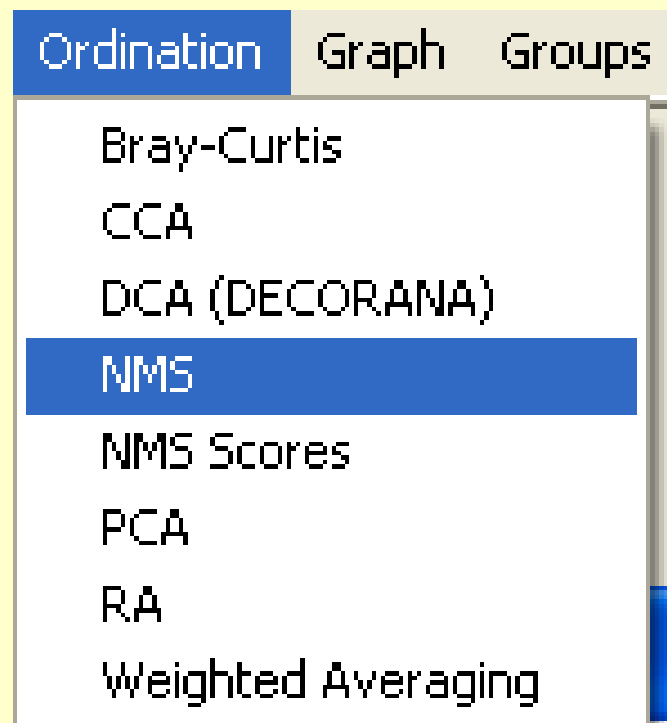
➤ *Objectives:*

- Showcase NMDS analysis – in PC-ORD and the literature

# NMS – Suggested Procedure

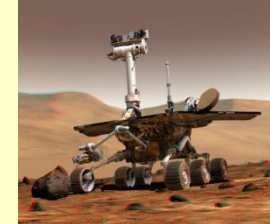
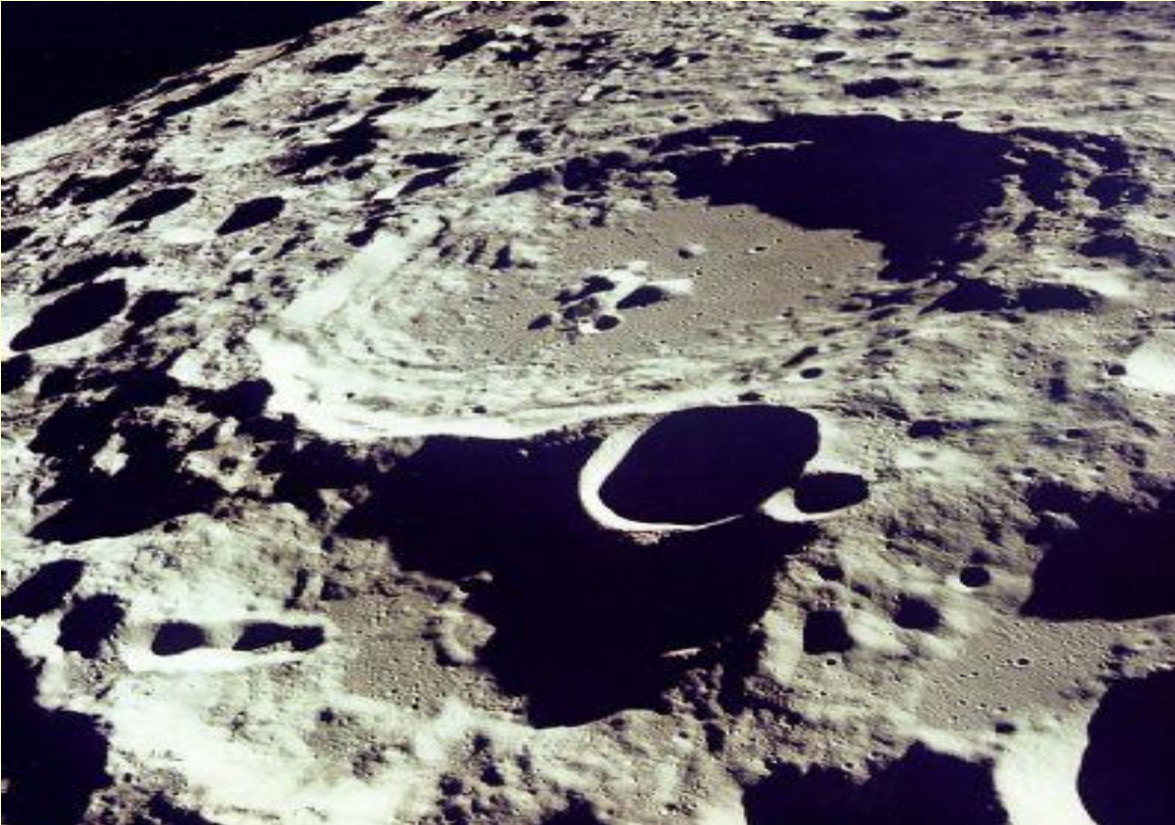
## (McCune and Grace 2002)

These “minimum” suggested procedures for determining appropriate largest dimensionality, assigning statistical significance with randomizations, and avoiding local minima.



- **Recommendation:** Request a 6-dimensional solution, stepping down to a 1-dimensional solution, with instability criterion of 0.0005, 200-500 iterations, 20-50 runs with real data, and 20-50 runs of randomized data  
**(NOTE: use more runs for randomization significance tests)**

# NMS – Suggested Procedure (McCune and Grace 2002)



Iterations:  
250 – 500

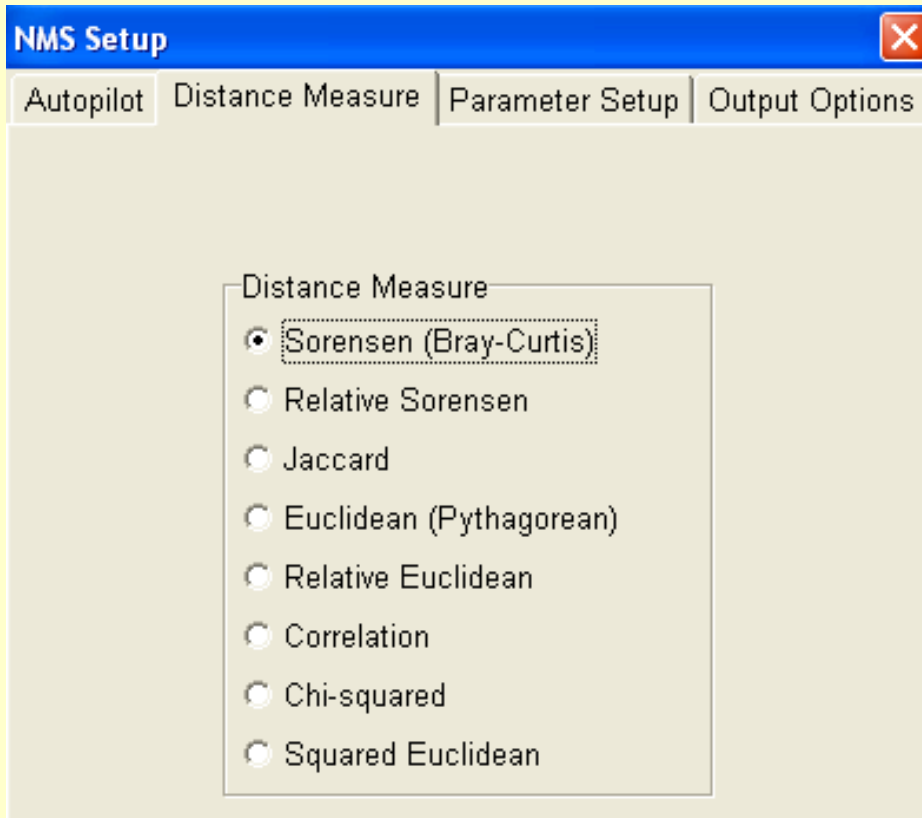
Robots  
per  
mission

Runs:  
20 – 50

Missions

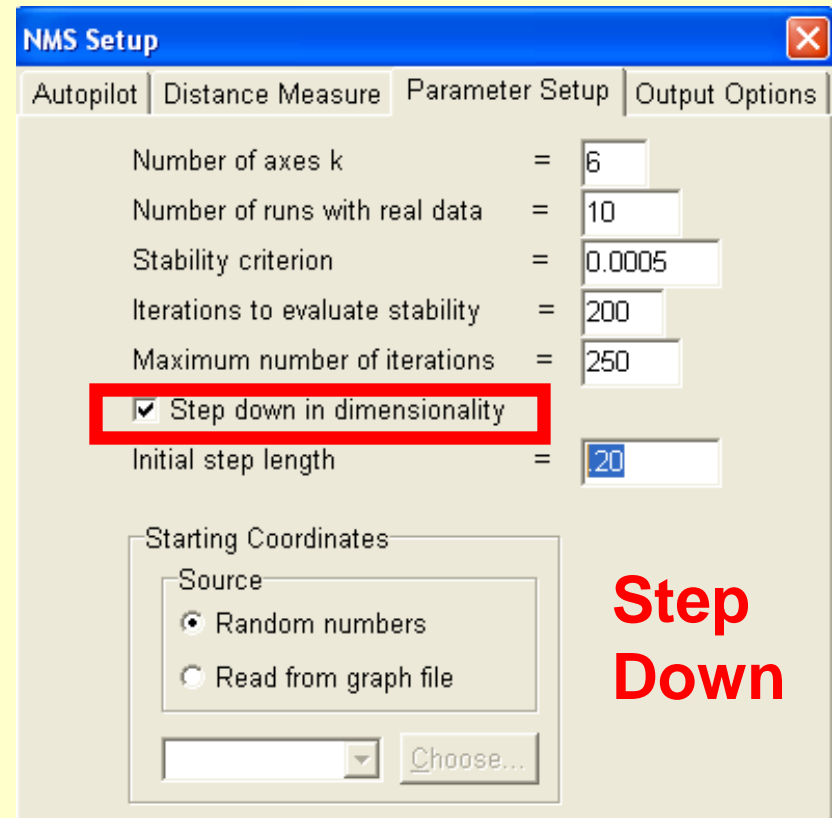
# NMS – Suggested Procedure: Step1

➤ First, pick distance measure



- Relative Sorensen
- Relative Euclidean

➤ Second, set up parameters



**Step  
Down**

- Dimensions (max = 6)
- Stepping Down

# NMS – Suggested Procedure

➤ Third, pick the output options

**NMS Setup** [X]

Autopilot | Distance Measure | Parameter Setup | **Output Options**

Write distance matrix

Write starting coordinates

List stress, etc. for each iteration

Write final configuration **Provides scores**

Plot distance vs dissimilarity

Randomization test

Varimax rotation

Run log **Statistics Dimensionality**

Plot stress vs. iteration

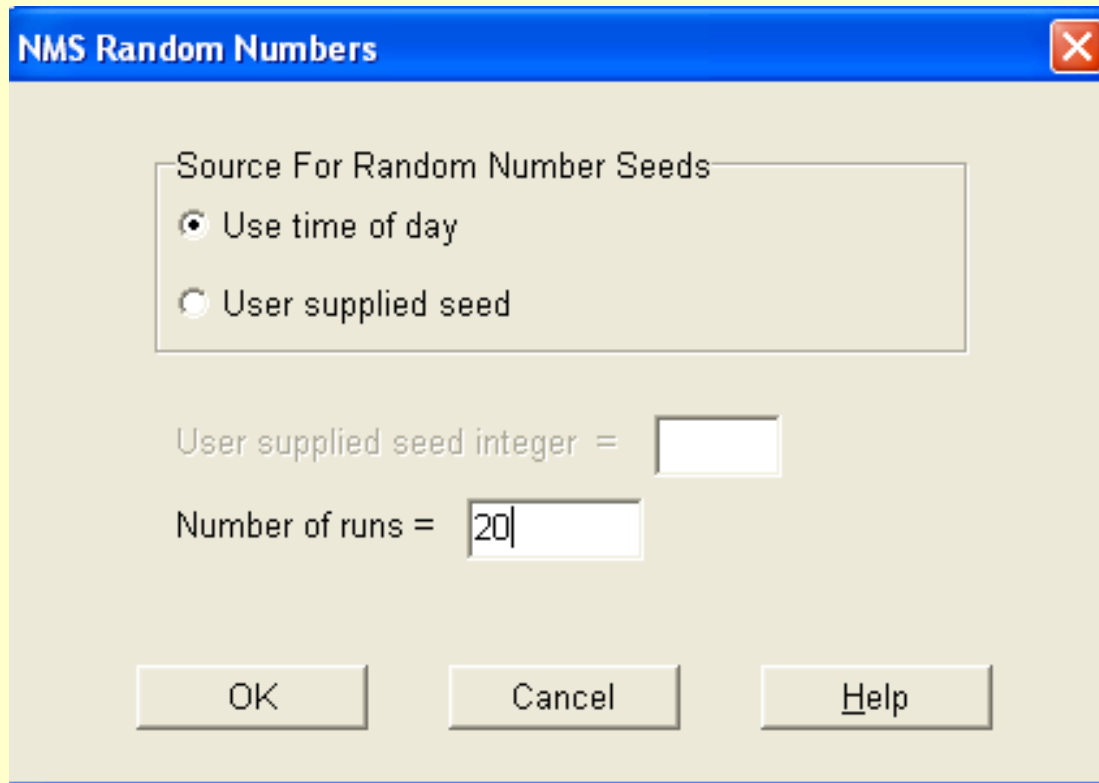
Calculate scores for species by weighted averaging

- Write final configuration
- Run Log
- Plot Stress vs.. Iteration
- Plot distance vs. dissimilarity
- Randomization Statistical Test
- Species Scores (for plotting)

# NMS – Suggested Procedure

## 1. Preliminary runs: Stress Test determines dimensionality

- Use “time of day” – random seed
- “Graph” messages



**NMS Random Numbers**

Source For Random Number Seeds

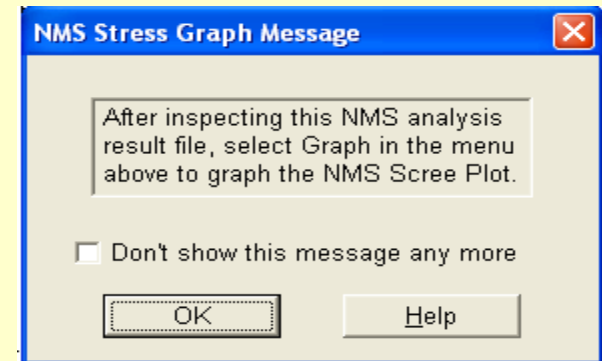
Use time of day

User supplied seed

User supplied seed integer =

Number of runs =

OK Cancel Help

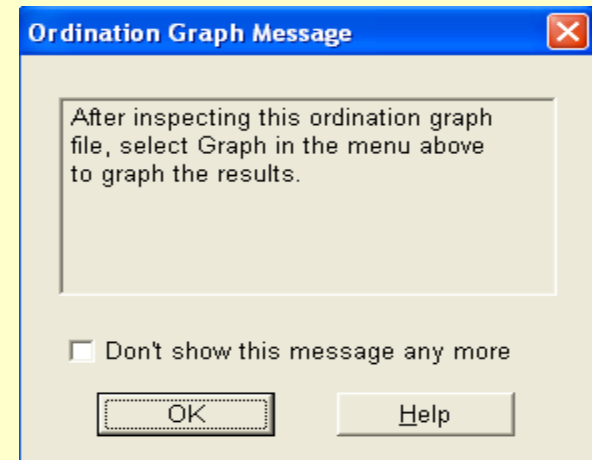


**NMS Stress Graph Message**

After inspecting this NMS analysis result file, select Graph in the menu above to graph the NMS Scree Plot.

Don't show this message any more

OK Help



**Ordination Graph Message**

After inspecting this ordination graph file, select Graph in the menu above to graph the results.

Don't show this message any more

OK Help

# 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.           250 = 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.           10 = Number of runs with real data
9.           20 = Number of runs with randomized data
10.          NO = Autopilot
11.  0.000500 = Stability criterion, standard deviations in stress
              over last 200 iterations.

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

-----
```

# NMS – Results

- Examine Results.txt file: Results for each run / dimension

```
***** 6-dimensional solution *****
```

```
4.13784 = final stress for 6-dimensional solution  
0.00181 = final instability  
250 = number of iterations
```

**Stress**

**Scores**

Final configuration (ordination scores) for this run

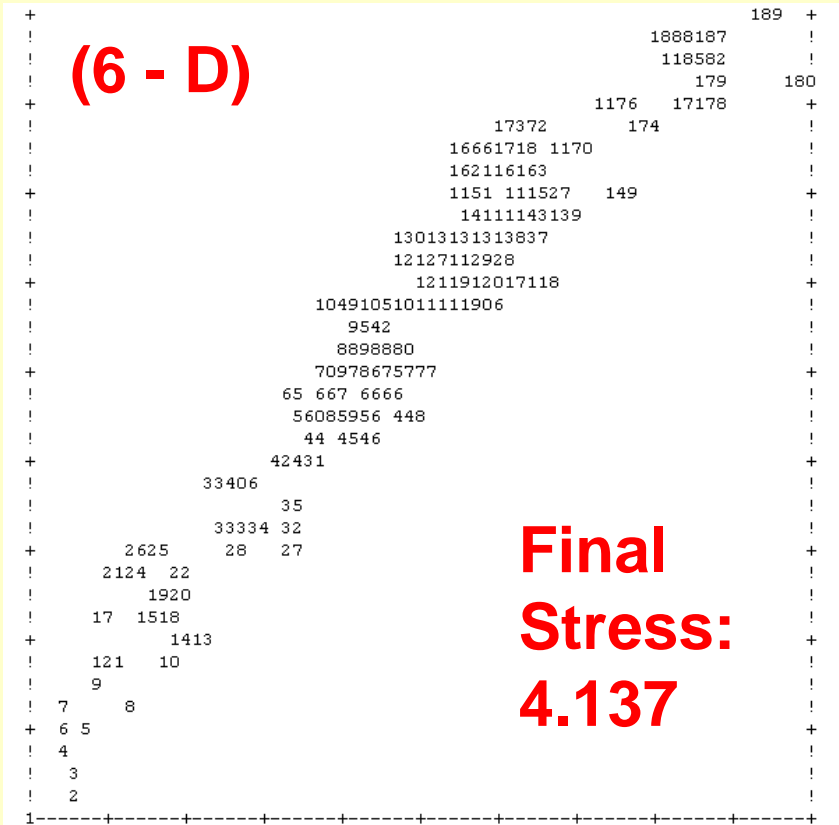
stands	Axis	1	2	3	4	5	6
Number Name							
1 Cst1		-0.4593	-0.4917	-0.1972	-0.4748	0.6636	-0.9756
2 Cst10		0.3223	0.3019	0.3186	-0.0754	0.2269	0.6306
3 Cst11		-0.2877	0.5412	-0.0581	0.1729	0.0283	0.5286
4 Cst13		0.3234	0.5688	0.6306	0.3683	0.1842	0.0530
5 Cst14		-0.1773	-0.2887	-0.0626	0.2415	0.9185	0.0570
6 Cst15		0.0429	0.1813	-0.6542	0.3843	-0.1038	-0.0570
7 Cst2		-0.1684	-0.4507	-0.0887	-0.1193	0.4672	-0.5257
8 Cst5		-0.4274	0.6304	-0.3393	-0.1422	0.1002	0.0534
9 Cst8		-0.0036	-0.3303	-0.1930	0.3138	0.1324	-0.1652
10 Cst9		0.5905	0.8432	0.6634	0.2364	0.4528	-0.4135
11 CscC		0.6282	0.2145	-0.5173	-0.6371	0.1019	-0.2152
12 CscD		-0.3524	-0.5933	-0.2223	0.0561	0.0939	0.1105
13 CscE		0.6514	-0.2372	-0.0369	0.6826	-0.3325	-0.2327
14 CscG		-0.3171	0.4538	0.5825	-0.1828	-0.7567	-0.5066
15 CscL		-0.4203	-0.5140	0.1037	-0.0092	-0.3429	0.7900
16 CscO		-0.5565	-0.3004	-0.0761	0.1069	-0.4352	0.0866
17 CscP		0.6217	-0.5315	0.0009	0.2558	-0.8028	-0.0575
18 CscQ		-0.1452	0.2822	0.2895	-0.7168	0.0039	0.2332
19 CscS		0.3446	-0.0388	0.3447	-0.7647	-0.1482	0.0244
20 CscT		-0.2097	-0.2407	-0.4884	0.3035	-0.4517	0.5819



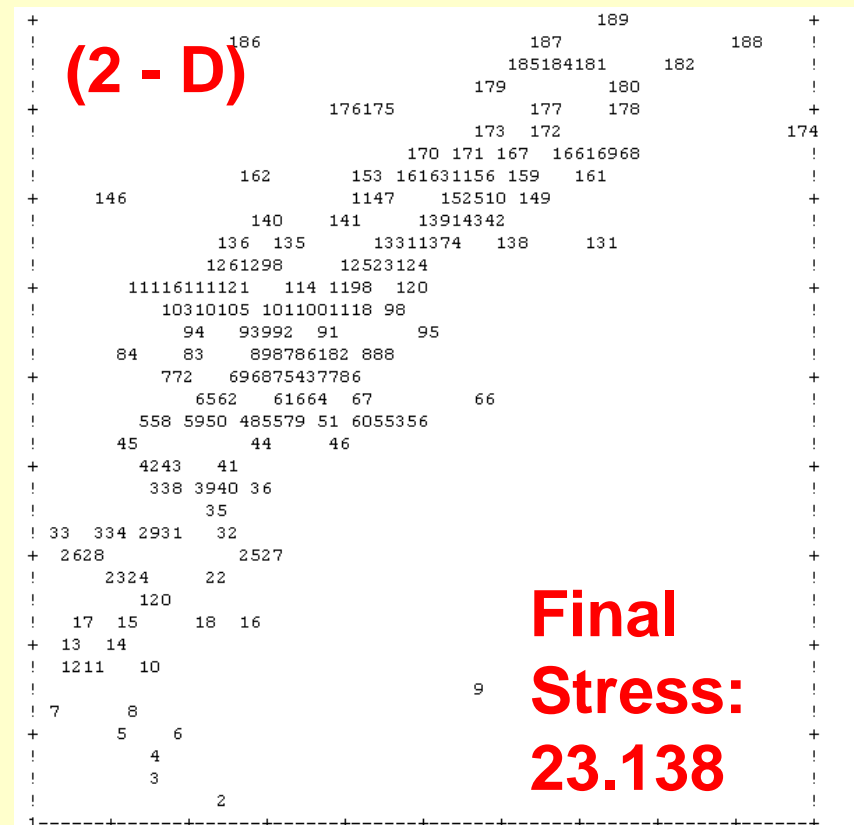
# NMS – Results

- Examine Results.txt file: Shepard Diagram

Differences in Real Space



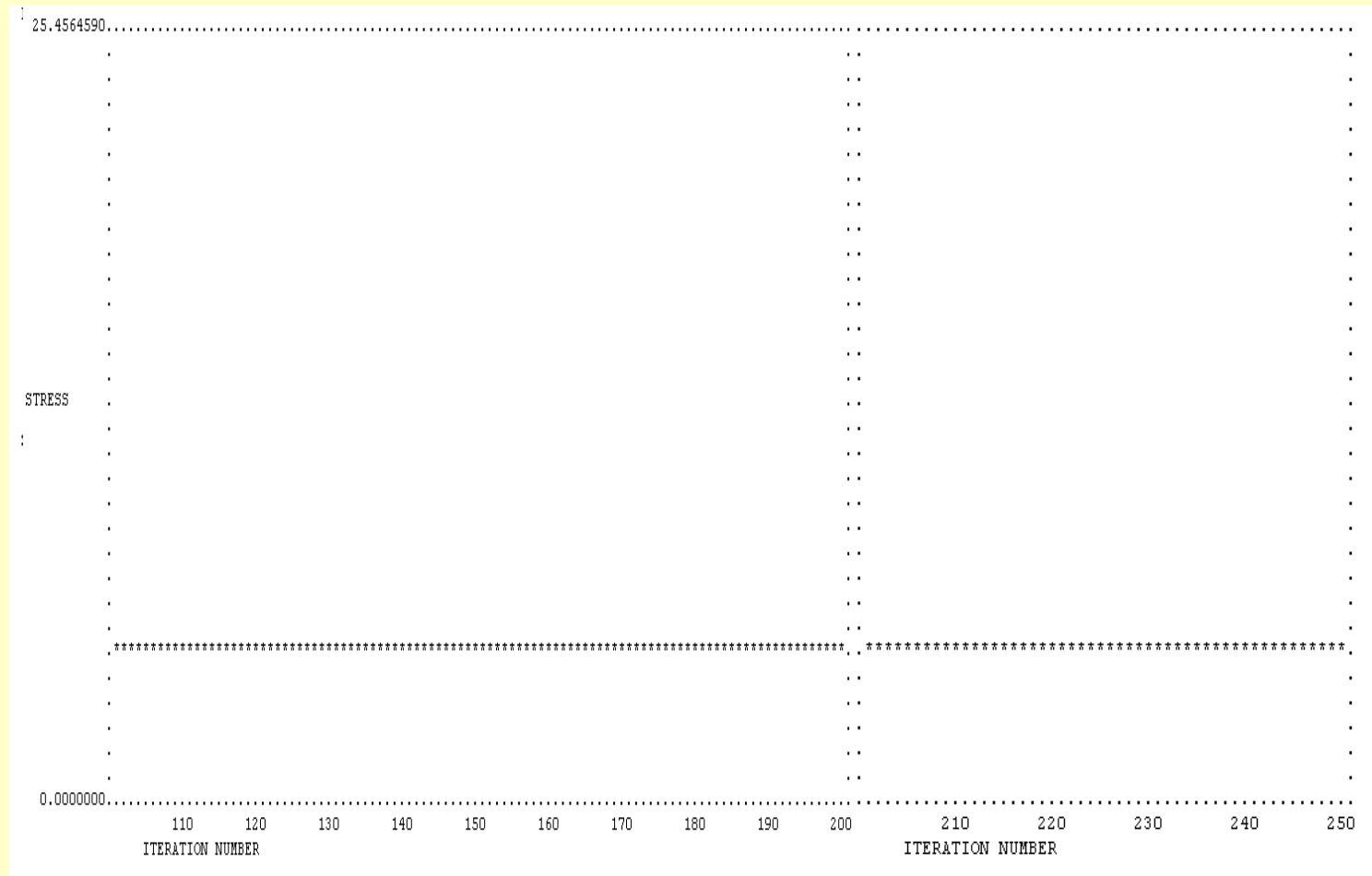
Distances in 6-D space



Distances in 2-D space

# NMS – Results

➤ Examine Results.txt file: Plotting Stress vs. Iteration



- **Note:** This process is repeated for each run

# NMS – Results

## ➤ Examine Results.txt file: Stress

STRESS IN RELATION TO DIMENSIONALITY (Number of Axes)

Axes	Stress in real data 10 run(s)			Stress in randomized data Monte Carlo test, 20 runs			p
	Minimum	Mean	Maximum	Minimum	Mean	Maximum	
1	38.376	46.541	54.222	41.561	48.626	54.483	0.0476
2	20.366	22.469	25.766	21.752	24.574	28.997	0.0476
3	13.418	13.670	14.855	13.809	15.954	17.877	0.0476
4	8.919	8.954	9.268	8.579	10.807	12.085	0.0952
5	6.078	6.288	6.587	6.662	7.863	9.987	0.0476
6	4.138	4.217	4.499	4.635	5.716	7.708	0.0476

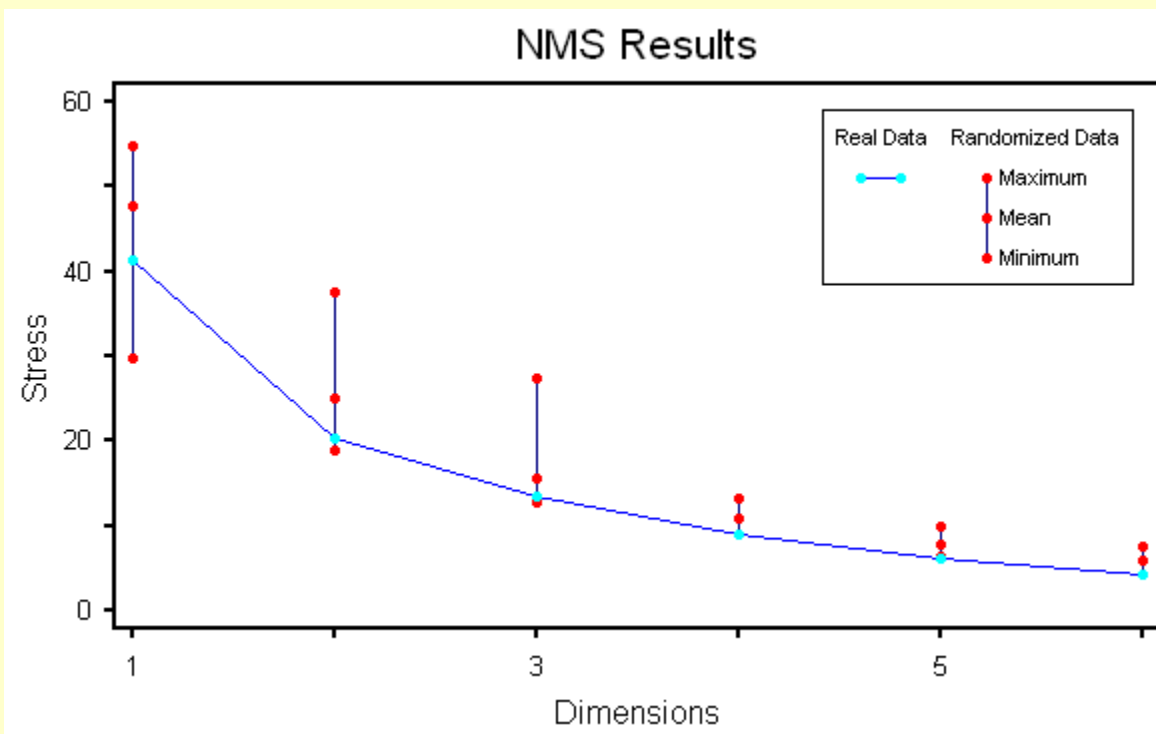
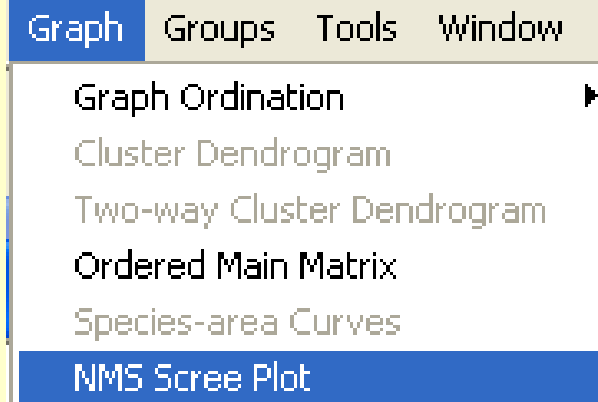
p = proportion of randomized runs with stress < or = observed stress  
i.e.,  $p = (1 + \text{no. permutations} \leq \text{observed}) / (1 + \text{no. permutations})$

Number of Axes – p values: 3 (stress = 13.418)

# NMS – Suggested Procedure: Step2

➤ **Goal:** Select the Best Solution:  
Plot stress vs. dimensions

**How:** After running NMS in PC-ORD  
Use: Graph | NMS Scree Plot



**NOTE:**

If the stress increases with additional dimensions, the model is over-fitted

# Information Theory - Suggested Procedure

- Trade-off between model fit and complexity (e.g., Akaike information criterion)

Measure of relative goodness of fit of a statistical model

Quantifies tradeoff between accuracy / complexity of model

$$AIC = 2k - 2 \ln(L)$$

- Where:

$k$  is the number of parameters in the statistical model

$L$  is the likelihood function for the estimated model

# NMDS model - Suggested Procedure

- Trade-off between model fit and complexity

Measure of relative goodness of fit of the NMDS

Quantifies tradeoff between accuracy / complexity of NMDS

$$NMDS\_Criterion = Stress + (Axes * K)$$

- Where:

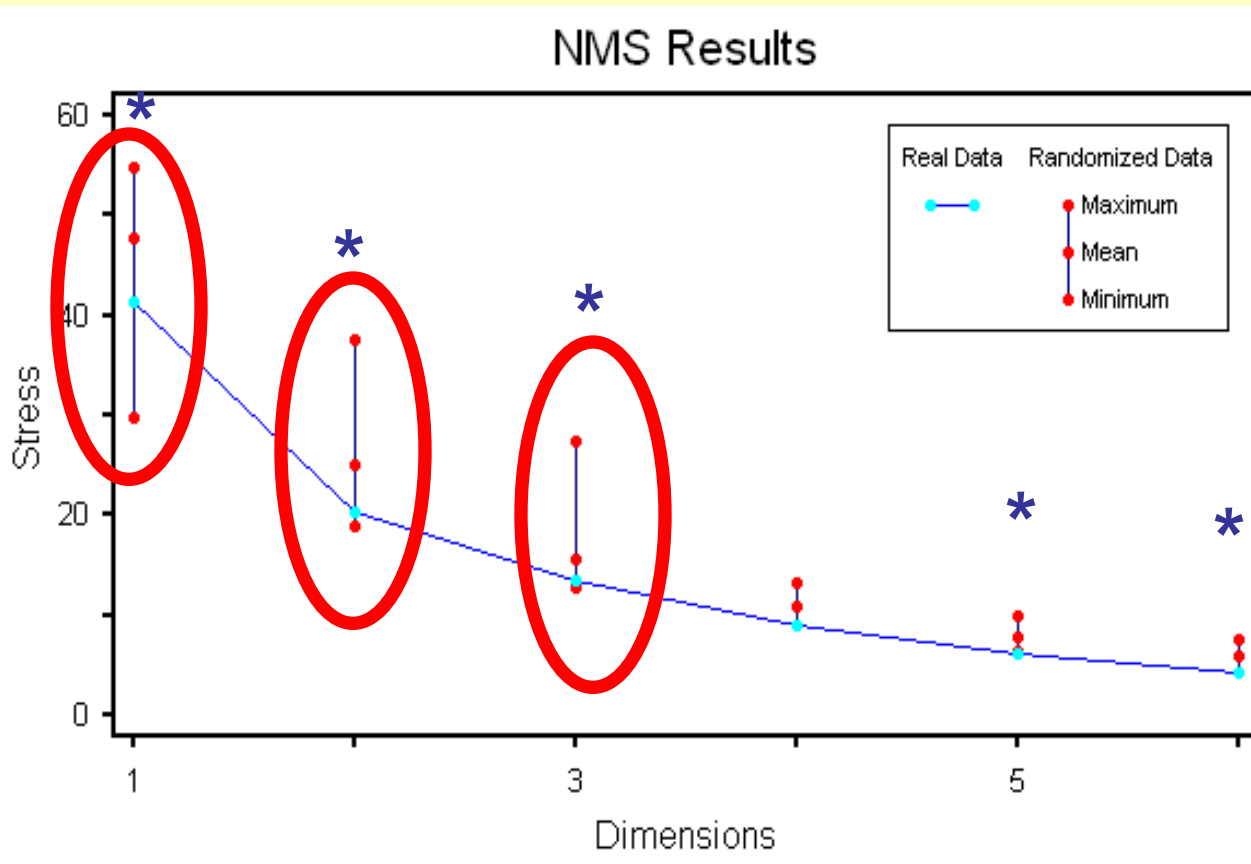
*Axes = number of axes (dimensions) in model*  
*K is the constant “stress” penalty for each axis*

# NMS – Suggested Procedure:

- PC-ORD uses the following criteria (for reference):
  - Comparing final (minimum) stress values among the best solutions, picks one best solution for each dimensionality.
  - Additional dimensions considered useful if they reduce final stress  $\geq 5$  (on a scale of 0-100). PC-ORD selects the highest dimensionality that meets this criterion.
  - At that dimensionality, final (minimum) stress must be lower than that for 95% of randomized runs (i.e.  $p \leq 0.05$ ).
  - If this criterion is not met, PC-ORD does not accept that solution and chooses a lower-dimensional solution, provided that it passes the specific randomization test.

# NMS – Suggested Procedure:

- **Criteria for selecting number of dimensions:**
  - marginal change in stress
  - p values



**NOTE:**

If stress does not increase, with added Ds, computer considers marginal decline with added Ds

Next, consider the p values



# NMS – Suggested Procedure:

➤ **Goal:** Check for a better-than-random solution using the results of the **Randomization test** (p values)

➤ **Limitations:** Helpful but not fool-proof

The most common problems are:

**Strong outliers, single super-abundant species, small data sets (e.g., < 10 SUs), many zeros**

➤ **Note:** The first axis with randomized community data is often as strong or stronger than the real data.

The randomization creates rows with unequal abundances. Thus the 1-D NMS solution from the shuffled data tends to capture variation in row totals.

**Interpret 1-D solutions carefully**

# NMS – Suggested Procedure:

- **Goal:** Select number of dimensions beyond which additional dimensions provide only small stress reductions
- **Suggestion:** Follow PC-ORD's recommendation – but check for safeguards
- **Note:**
  - “No firm fixed criterion for selecting an appropriate number of dimensions” (Kruskal and Wish 1978)
  - Axis scores depend on the number of axes.
  - Remember: First dimension on best multi-Dimensional (2-D, 3-D, ...) result will be different

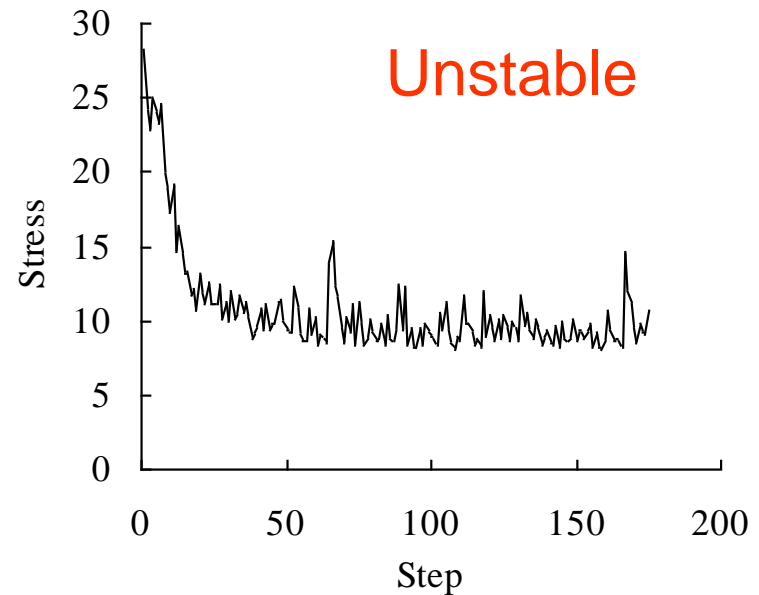
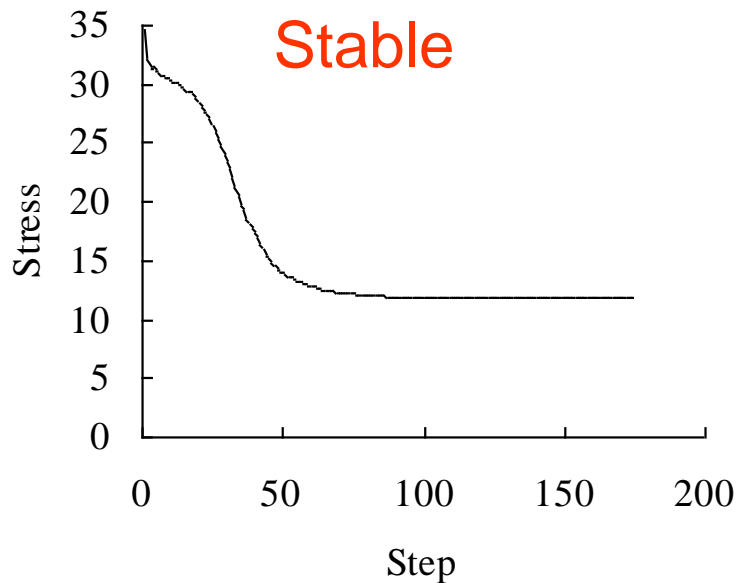
# NMS – Suggested Procedure:

## ➤ Trade-Offs:

- Final stress decreases and the proportion of the variance represented increases with more axes
- Complexity of explanation increases with more axes
- Pick as few (significant) dimensions as possible
  - based on stress reductions
- Beware of unstable results (stress wiggles with iterations)
- Consult the instability of the final answer
- NO matter what:
  - Do not trust results with large stress values ( $> 20$ )

# NMS – Suggested Procedure:

- Check the following plot of stress vs. iteration for stability for the NMDS chosen solution
  - Look for smooth curves
  - Strive for instability  $< 10^{-3}$  ( $< 0.001$ )

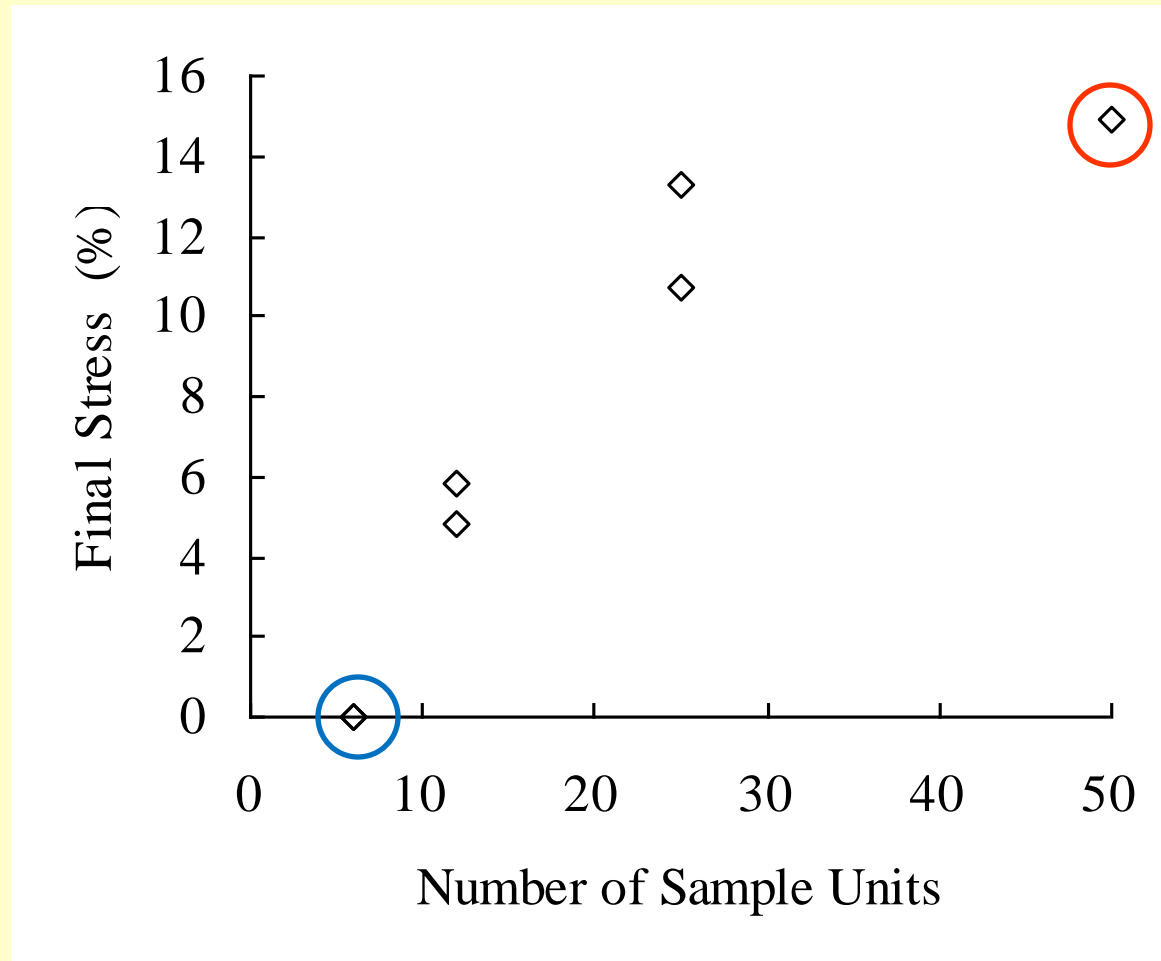


# NMS – Suggested Procedure:

- **Use Data Exploration** to explore stress of NMS analysis

Assessing the dependence of stress on sample size, by subsampling rows of matrix of 50 samples by 29 species

**Fewer samples lead to lower stress**

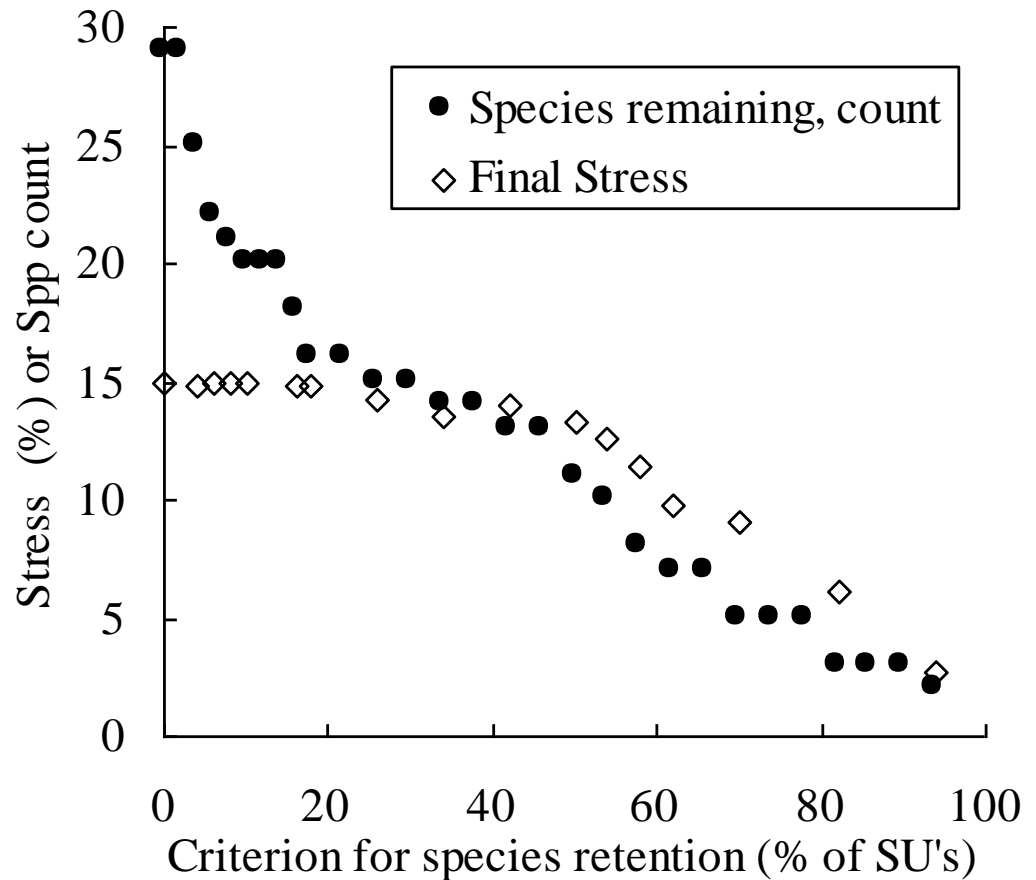


# NMS – Suggested Procedure:

- **Use Data Exploration** to explore stress of NMS analysis

Assessing the dependence of NMDS stress on progressive removal of rare species from the data set

Fewer species lead to lower stress



# NMS – What to Report

- Samples / Species Considered
- Data Transformations / Relativizations
- Distance measure used
- Did you use a random starting point ?
- Number of runs with real / random data
- Number of dimensions considered
- How did you select the dimensions
- Final stress / instability of best solution
- Monte Carlo tests results (runs, p values)
- Proportion of variance explained by each axis ( $r^2$ )
- Plot Overlays (env. data / species)
- Correlations of env. data / species with axes (Tau)

# NMS – References

- PC-ORD uses the following algorithms:

Mather, P. M. 1976. Computational methods of multivariate analysis in physical geography. J. Wiley & Sons, London. 532 pp.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonnumeric hypothesis. *Psychometrical* 29: 1-27.

- For a review of NMS, cite:

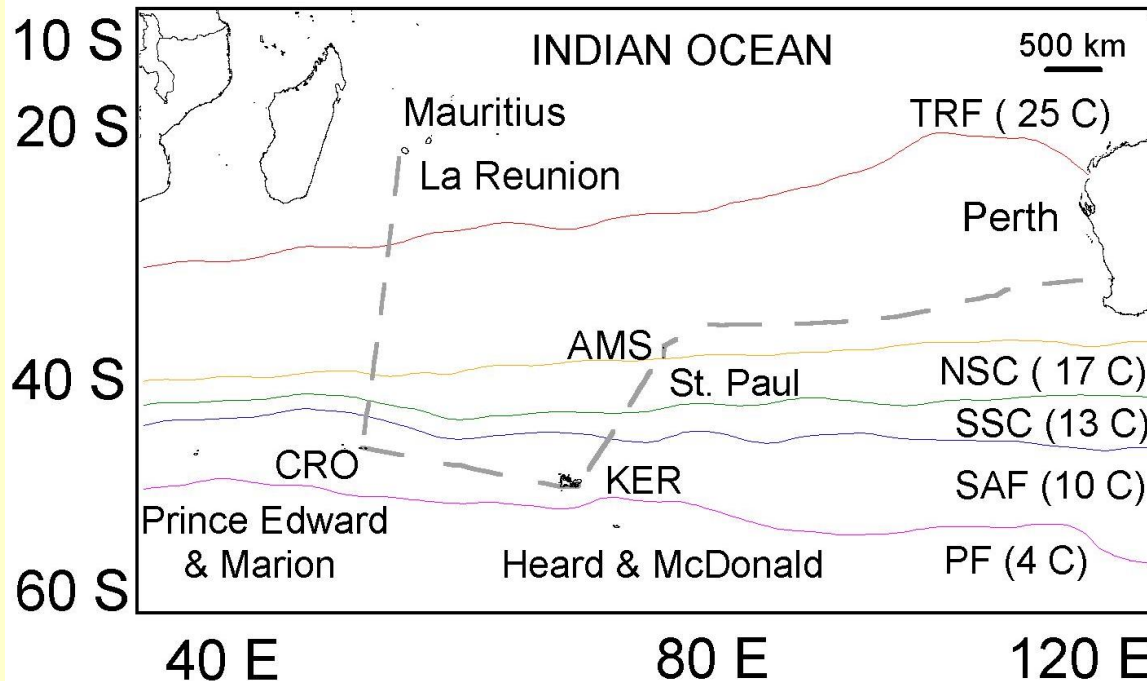
Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18: 117-143.

Kneel, N.C., Orloci, L., 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology* 67: 919–923.



# NMS – Examples I

## ➤ Seabird communities of the Indian Ocean



Selected an observation day as the sampling unit for the community level analysis because we regarded the daily transects as discrete samples, separated by night time periods with no survey effort.

**(Hyrenbach et al. 2007)**

Our sample size was a matrix of 16 daily transects and 46 taxa. We standardized data using relative abundance (birds / km<sup>2</sup>) of the taxa. To ensure each sample was weighted equally in the analysis, we used the relative Sorensen (Bray-Curtis) distance measure (Manly, 1994).

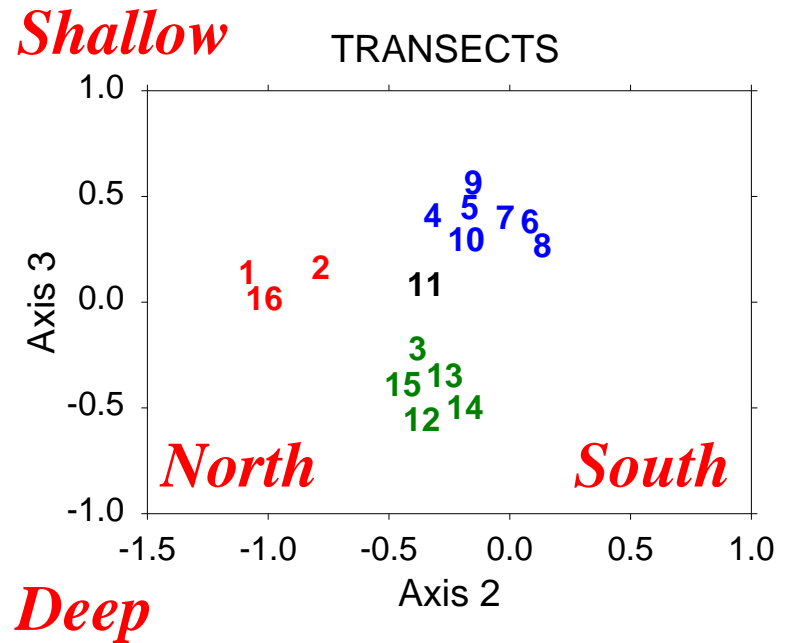
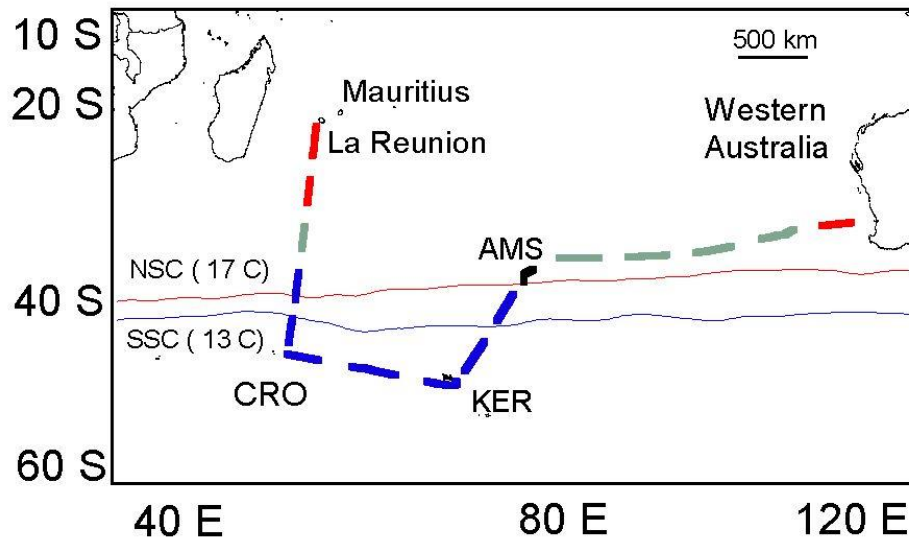
# NMS – Examples I

- Seabird communities of the Indian Ocean
  - The NMS selected 3 habitat axes, which accounted for 73.4 % of variance observed in the seabird community
    - The first axis ( $R^2 = 0.15$ ) described latitudinal gradients associated with a concurrent SST decrease and CHL increase (to the south).
    - The second axis ( $R^2 = 0.41$ ) illustrated concurrent lat / long changes in wind speed, depth, CHL, SST, and gradients in ocean depth and SST.
    - The third axis ( $R^2 = 0.17$ ) captured the influence of onshore–offshore gradients in CHL, irrespective of lat and long.
  - Because axis 2 and 3 explained a higher proportion of the observed variability, we plotted the survey transects and species distributions in 2-dimensions

(Hyrenbach et al. 2007)

# NMS – Examples I

- Seabird community structure in the Indian Ocean



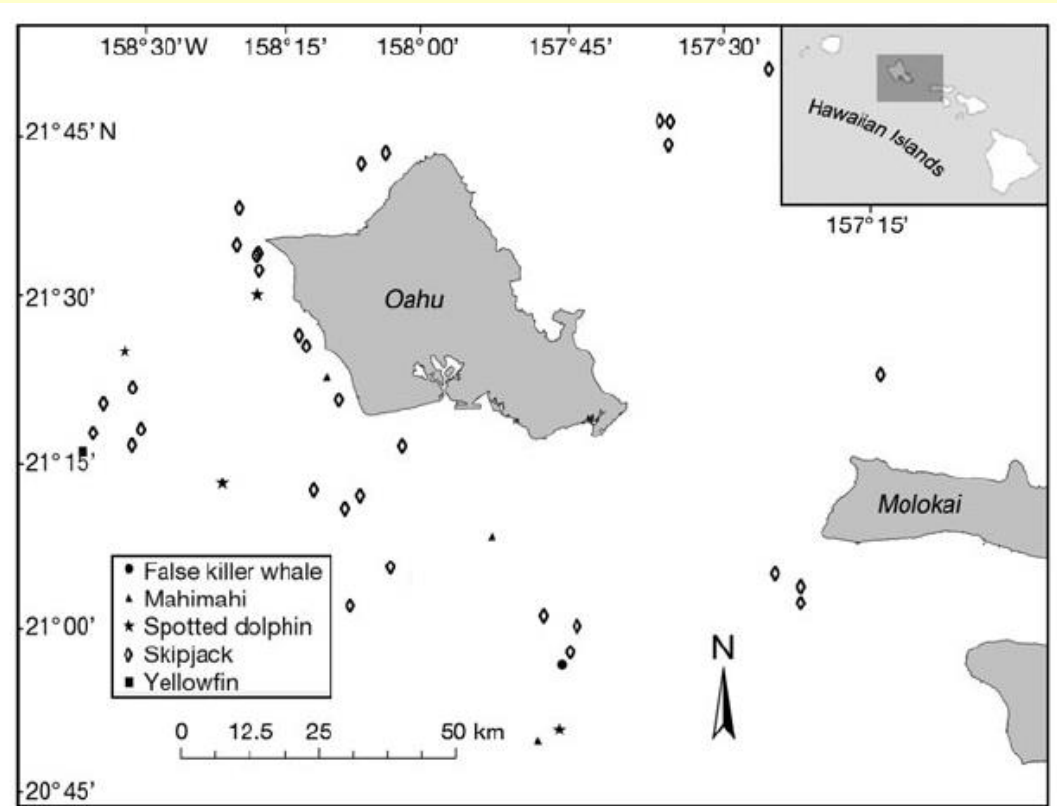
Three seabird assemblages:

sub-Antarctic, subtropical offshore, subtropical nearshore

(Hyrenbach et al. 2007)

# NMS – Examples II

## ➤ Seabirds and subsurface predators around O'ahu



69 seabird foraging observations recorded

Presence of subsurface predators was not ascertained in 7 schools

In 2 of 62 remaining observations, no subsurface predators were present

(Hebshi et al. 2008)

# NMS – Examples II

## ➤ Seabirds and subsurface predators around O'ahu

The NMDS analysis relied on a similarity matrix created using the Sorensen (Bray-Curtis) index from the raw seabird counts and 13 explanatory variables describing:

- type of fishing (commercial vs. sport)
- subsurface predator (skipjack tuna, mahimahi, spotted dolphin, false killer whale, yellowfin tuna, unknown),
- geographic location around O'ahu  
(Waianae, Penguin Bank, Kaena Point, other \*).

**\* Only those locations contributing at least 10%  
(7 or more) observations considered in analysis.**

**(Hebshi et al. 2008)**

# NMS – Examples II

## ➤ Seabirds and subsurface predators around O’ahu

NMS identified 2 highly (99.3%) orthogonal axes ( $r = -0.082$ ), which explained 67.9% of the cumulative observed variance

axis1,  $R^2: 0.502$

axis2:  $R^2: 0.178$

But NMDS stress was high (17.873), suggesting that the test performance was “fair”, and the results should be interpreted with caution (McCune & Grace 2002)

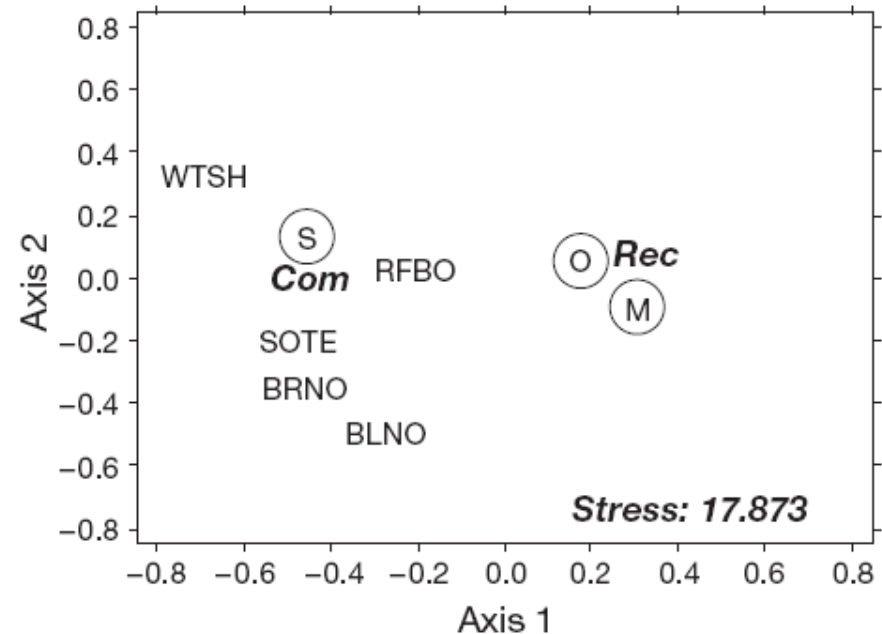


Fig. 7. NMDS plot showing 2-dimensional distances among the 5 most abundant seabirds (WTSH = wedge-tailed shearwater, SOTE = sooty tern, RFBO = red-footed booby, BLNO = black noddy, BRNO = brown noddy), 2 fishery types (commercial [Com] vs. recreational [Rec]), and 3 subsurface-predator types (M = mahimahi, O = odontocete, S = skipjack)

(Hebshi et al. 2008)

# NMS – Examples II

## ➤ Seabirds and subsurface predators around O’ahu:

- The seabird community was influenced by the presence of 3 species: wedge-tailed shearwaters, brown noddies and sooty terns
- The first axis captured the differences between commercial and sport fishing vessels, while the second axis captured variability across geographic locations
- This analysis also revealed significant correlations with first axis for 2 subsurface predators: mahimahi (+) and skipjack tuna (-)

Table 2. Kendall correlations with first and second ordination axes (n = 67). Significant values are in bold. Significance levels are:  $\alpha_{0.005} = 0.341$ ,  $\alpha_{0.05} = 0.241$ . See Table 1 for scientific names of seabird species and predator type

	Axis 1 tau	Axis 2 tau
<b>Seabird species</b>		
Wedge-tailed sheawater	<b>-0.779</b>	<b>0.427</b>
Brown noddy	<b>-0.242</b>	<b>-0.537</b>
Black noddy	-0.168	-0.103
Brown booby	-0.037	-0.144
Red-footed booby	-0.093	-0.017
Sooty tern	<b>-0.106</b>	<b>-0.329</b>
White tern	-0.007	0.119
Sooty or short-tailed shearwaters	-0.089	-0.147
Pomarine jaeger	-0.104	-0.098
South polar skua	-0.163	0.201
Newell’s shearwater	-0.237	0.237
Great frigatebird	0.059	0.091
<b>Vessel type</b>		
Sport	<b>0.351</b>	<b>-0.011</b>
Commercial	<b>-0.351</b>	<b>0.011</b>
<b>Location</b>		
Waianae	<b>-0.078</b>	<b>0.315</b>
Penguin Banks	0.050	0.017
Kaena	-0.006	-0.125
Windward	<b>-0.008</b>	<b>-0.288</b>
Other	0.080	-0.125
<b>Predator type</b>		
Spotted dolphin	0.185	0.040
Mahimahi	<b>0.281</b>	<b>-0.070</b>
False killer whale	0.141	-0.016
Skipjack	<b>-0.452</b>	<b>0.134</b>
Yellowfin	-0.037	0.094
Unknown	0.241	-0.199

(Hebshi et al. 2008)

# Take Home Messages

- NMDS is a flexible and powerful tool
- NMDS computational approach allows the integration of different datasets into multivariate patterns
- This inherent flexibility makes this technique difficult to interpret due to the inherent variability (not deterministic)
- Use NMDS to tell ecological stories that balance the “noise” in the data with “statistical significance” of patterns.
- Mind the trade-off between simplicity and completeness.
- Data exploration can help you use NMDS most efficiently, by carefully choosing the sample sizes and species / variables to include in your analyses.