A 'fuzzy clustering' approach to conceptual confusion: how to classify natural ecological associations

Dario Fiorentino*, Roland Pesch, Carmen-Pia Guenther, Lars Gutow, Jan Holstein, Jennifer Dannheim, Brigitte Ebbe, Tim Bildstein, Winfried Schroeder, Bastian Schuchardt, Thomas Brey, Karen Helen Wiltshire

*Corresponding author: dario.fiorentino@awi.de

Content

Figure S1. Principal coordinate ordination plot.

Figure S2. Comparison between communities' distribution in the German Bight.

Figure S3. Maps of stations subsets with increasing equitability for each group.

Figure S4. Maps of stations subsets with increasing distance among station and constant number of stations.

Figure S5. Maps of stations subsets with changing amount of information.

Table S1. Monte Carlo Permutation test with the Pseudo-F of the distribution of beta diversity across the four communities.

Table S2. Monte Carlo Permutation test with the Pseudo-F of the distribution of beta diversity across three degrees of equitability separately for each community.

Table S3. Comparison between taxonomic ranks.

Table S4. Analysis of variance on confusion index distributions.
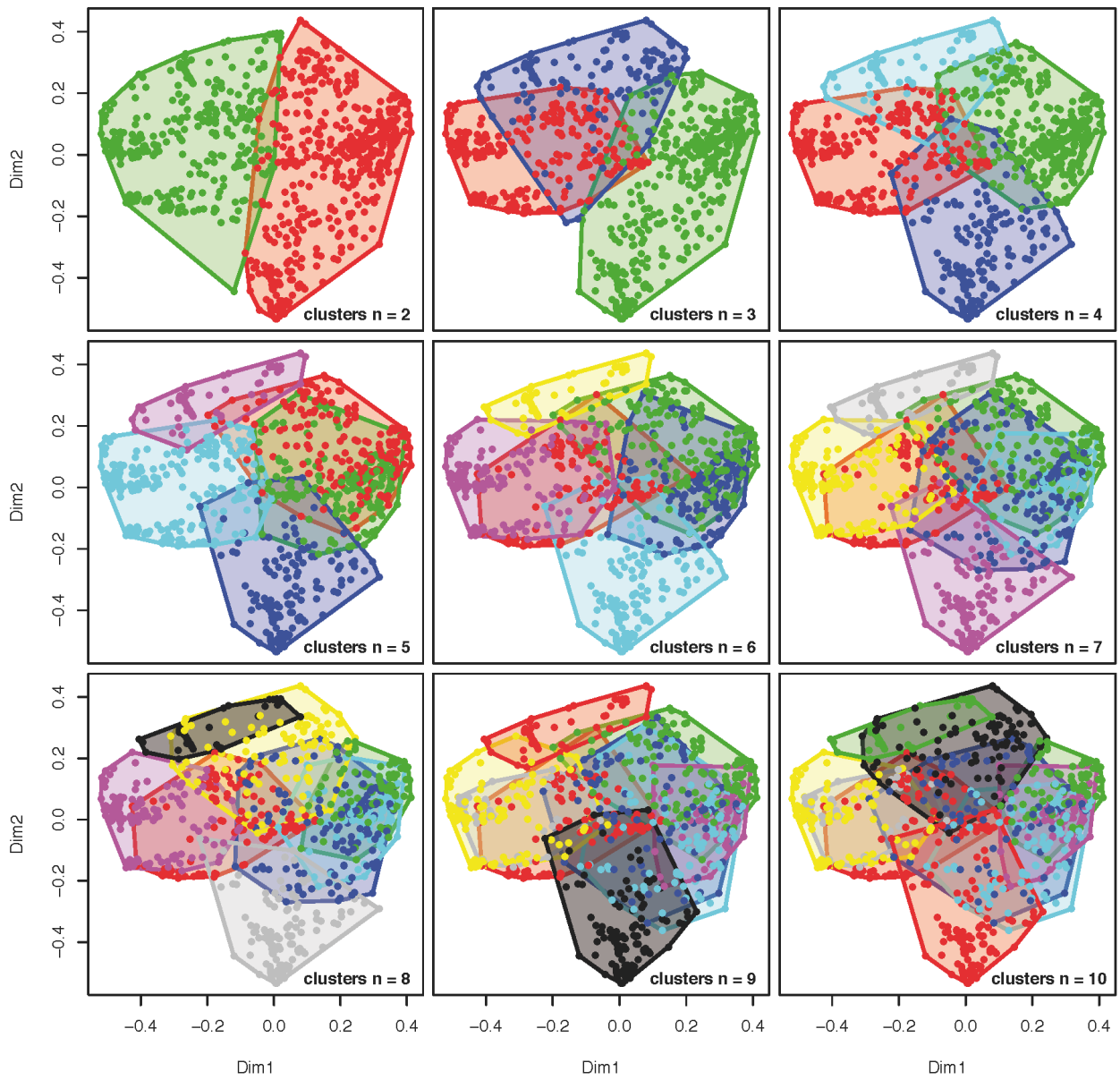
Literature cited

Fig S1. Principal coordinate ordination plot. This was used to visually find the optimal number of clusters and validate the classification in terms of separation and compactness. The ordination was run on Bray-Curtis dissimilarity of abundance data. Colours and hulls help to identify the different clusters.
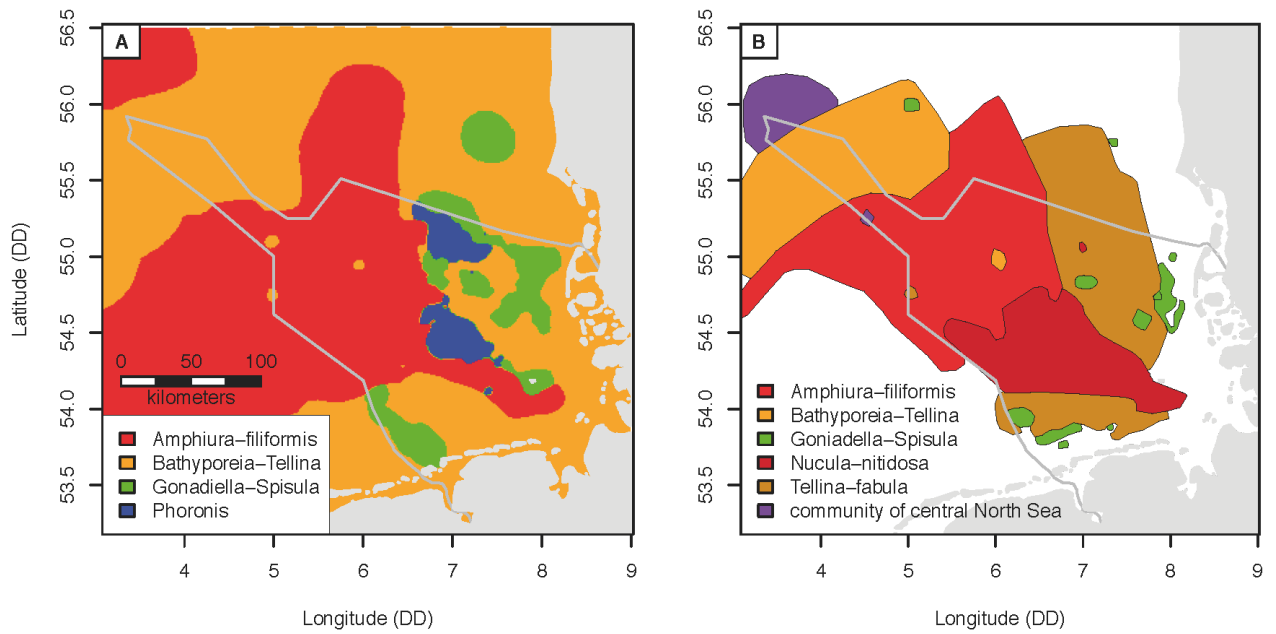
Fig S2. Comparison between communities' distribution in the German Bight. Panel A shows communities distribution found in the present study using crisp boundaries. Panel B shows the map provided by Rachor & Nehmer (2003). This was drawn by hand. Here we modified its original colouring in the way that those communities found both in the present study and by Rachor & Nehmer (2003) are sharing same colour shades. Maps were created with R version 3.2.3 (R Development Core Team 2015).
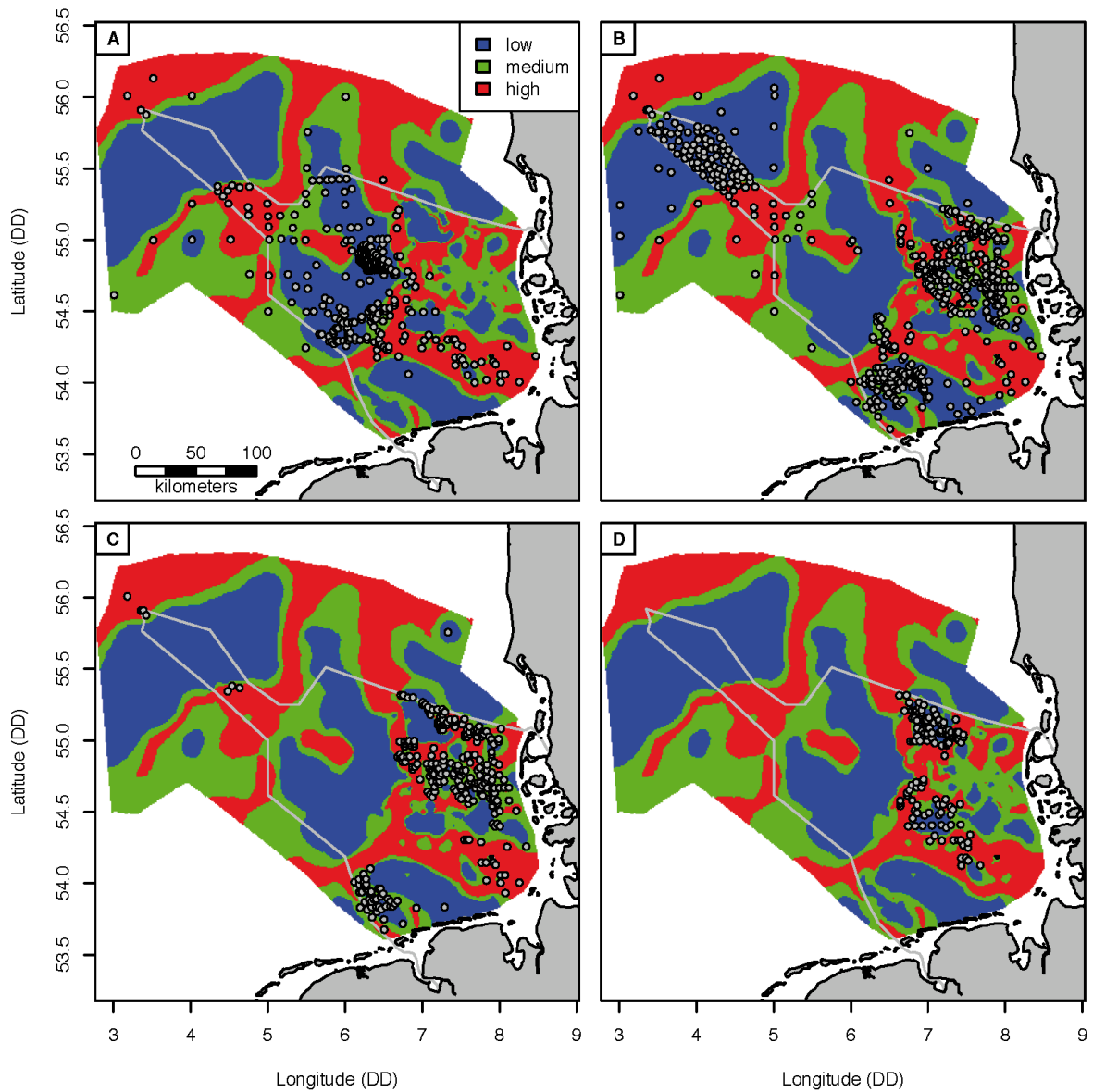
Fig S3. Maps of stations subsets with increasing equitability for each group. Panels show stations selected for groups (A) AMPHIURA-FILIFORMIS, (B) BATHYPOREIA-TELLINA, (C) GONIADELLA-SPISULA, (D) PHORONIS. Coloured areas show different degrees of equitability. Blue indicates low equitability (≤ 0.4), green medium (> 0.4 ≤ 0.6) and red high equitability (> 0.6). Maps were created with R version 3.2.3 (R Development Core Team 2015).
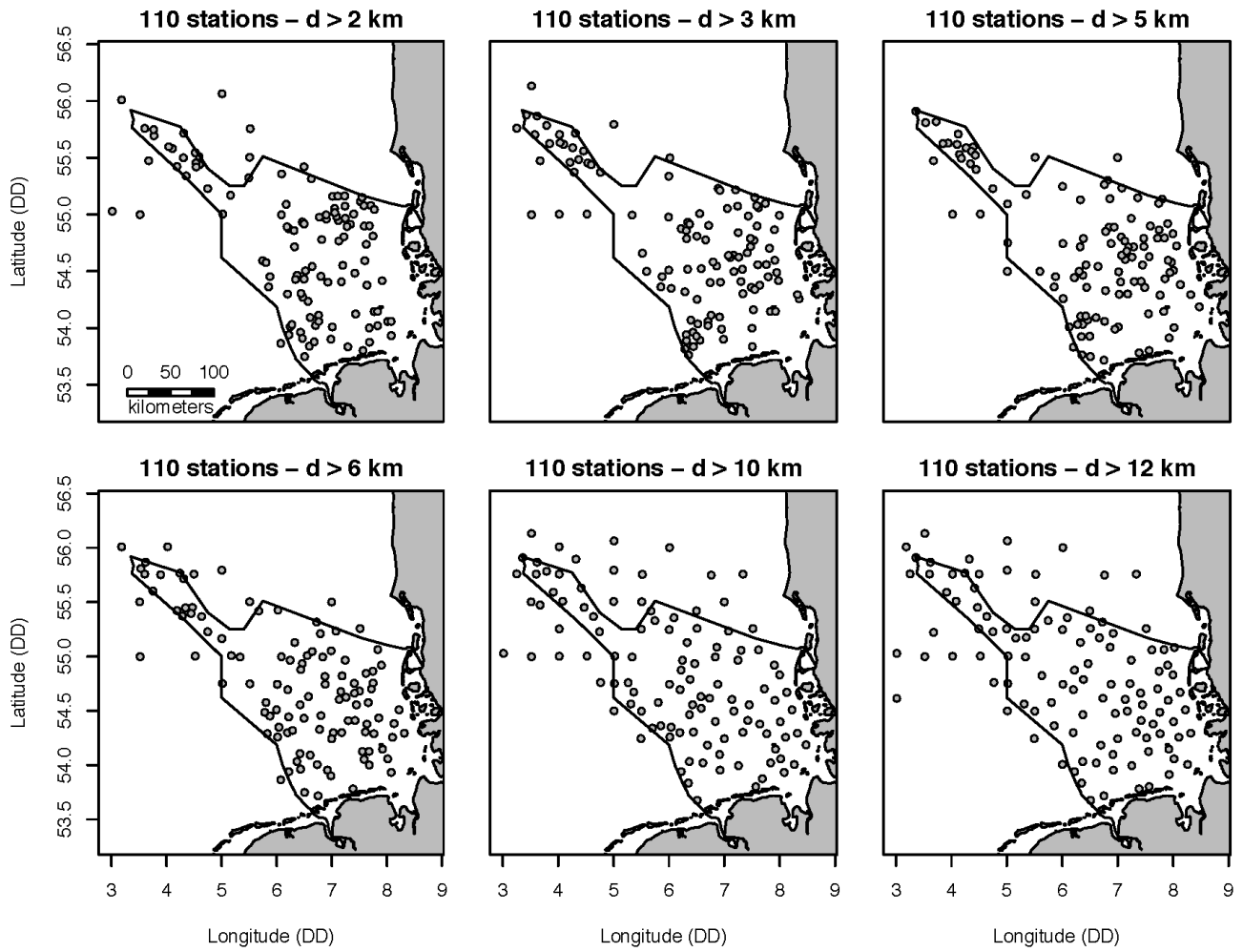
Fig S4. Maps of stations subsets with increasing distance among station and constant number of stations. The six different subsets show spatial pattern: from the smallest scale to the largest, the distribution of station changes from aggregated to even. Maps were created with R version 3.2.3 (R Development Core Team 2015).
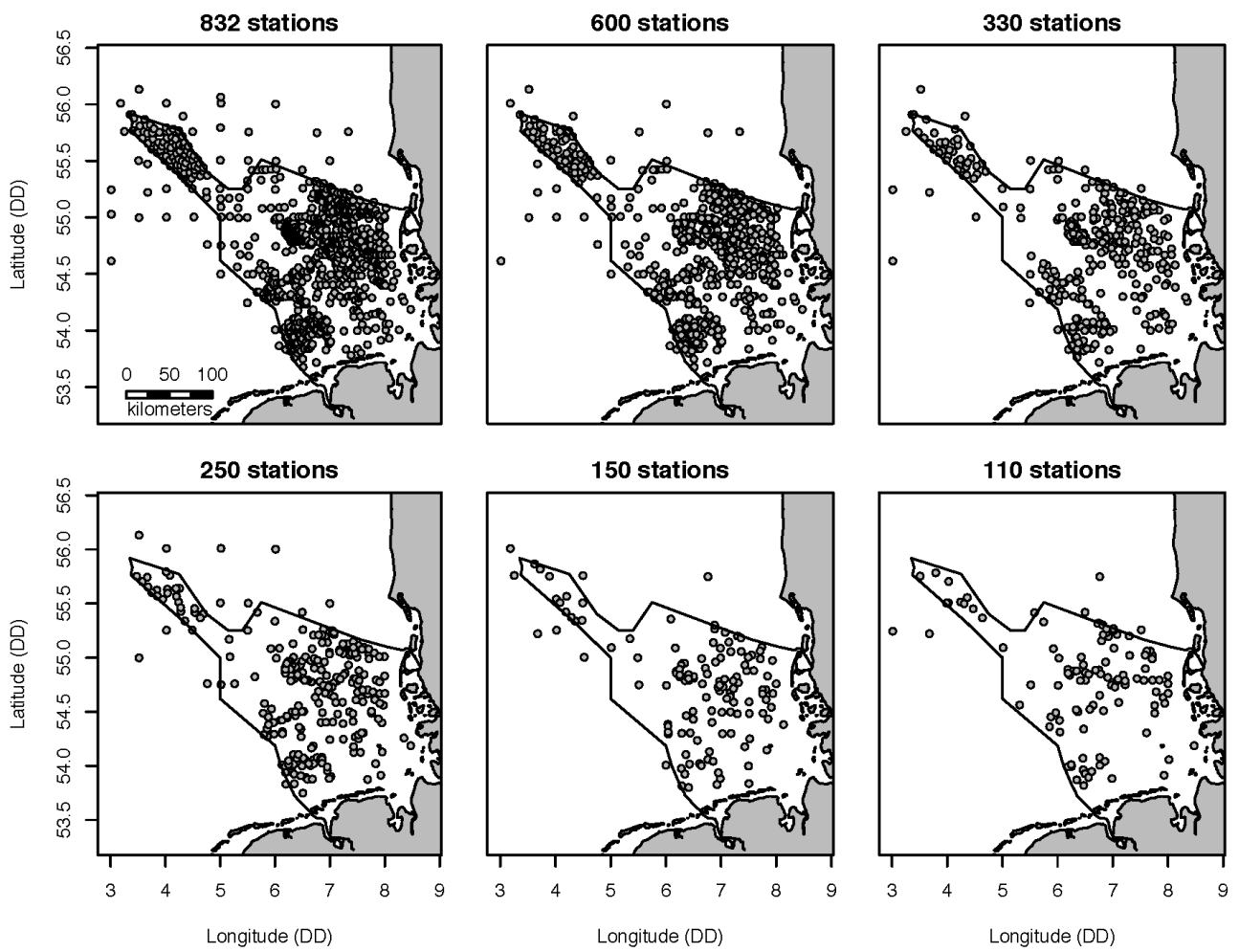
Fig S5. Maps of stations subsets with changing amount of information. The six different subsets show no spatial pattern, however the ratio between very densely sampled area and sparsely sampled areas is constant in all the subsets. Maps were created with R version 3.2.3 (R Development Core Team 2015).

Table S1. Monte Carlo Permutation test with the Pseudo-F of the distribution of beta diversity across **the four communities.** It was calculated according to the PERMDISP analysis performed on the Jaccard distance matrix based on species presence absence. Letters from "a" to "d" indicate the communities AMPHIURA-FILIFORMIS, BATHYPOREIA-TELLINA, GONIADELLA-SPISULA, PHORONIS respectively. ***, p ≤ 0.001

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Communities	3	2.38	0.792	156	1e-03 ***
Residuals	828	4.20	0.005		
Pairwise test	a ≠ b ≠ c ≠ d				

Table S2. Monte Carlo Permutation test with the Pseudo-F of the distribution of beta diversity across three degrees of equitability separately for each community. It was calculated according to the PERMDISP analysis performed on the Jaccard distance matrix calculated on species presence absence. Low indicates equitability ≤ 0.4, medium, equitability > 0.4 and ≤ 0.6, high, equitability > 0.6.

AMPHIURA-FILIFORMIS

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Degree of equitability	2	0.365	0.1825	32.6	0.001 ***
Residuals	288	1.612	0.0056		
Pairwise test	low ≠ medium = high				

BATHYPOREIA-TELLINA

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Degree of equitability	2	0.71	0.356	46.5	0.001 ***
Residuals	475	3.64	0.008		
Pairwise test	low ≠ medium = high				

GONIADELLA-SPISULA

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Degree of equitability	2	0.133	0.0667	20.9	0.001 ***
Residuals	265	0.848	0.0032		
Pairwise test	low ≠ medium ≠ high				

PHORONIS

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Degree of equitability	2	0.499	0.2493	9.99	0.001 ***
Residuals	121	3.019	0.0249		
Pairwise test	low ≠ medium = high				

Table S3. Comparison between taxonomic ranks.
Separately for each taxonomic rank, we tested the null hypothesis the number of species, genera, families or order does not change across the increasing gradient of equitability. For each cluster we fitted a generalized linear model assuming the data following a Poisson distribution. Terms were further tested with an analysis of deviance using a chi-square test. P-values are given. Red and blue fonts in bold indicate whether the significant variation with increasing degree of equitability was positive (red) or negative (blue).

		Taxonomic ranks			
		Species	Genus	Family	Order
Groups	Amphiura-filiformis	0.4456	**1.13e-14**	**2.641e-08**	**0.002066**
	Bathyporeia-Tellina	**0.006134**	0.7658	**3.169e-06**	**0.003297**
	Goniadella-spisula	0.2822	**0.001071**	**9.07e-05**	0.6956
	Phoronis	**1.124e-14**	**1.546e-07**	**0.0001928**	**3.539e-13**

Table S4. Analysis of variance on confusion index distributions. Analysis of variance for the linear model fitting the distributions of confusion index either across a) the 6 conditions simulating the effect of pure changing information or b) the 6 conditions simulating the effect of pure changing of spatial scale. Results were passed to a Turkey test (confidence level at 95%) for pairwise comparison between conditions. In bold p-values $\ll 0.001$.

a) Change of information

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Number of stations	5	0.2736	0.05472	136.2	**< 2e-16**
Residuals	2994	1.2031	0.00040		

Tukey Honest Significant Difference

	600	330	250	150	110
832	0.80	**0.01**	**0.00**	**0.00**	**0.00**
660		0.32	**0.00**	**0.00**	**0.00**
330			0.34	**0.00**	**0.00**
250				**0.00**	**0.00**
150					**0.00**

b) Change of spatial scale

Source	DF	Sum Sq	Means Sq	F	Pr (>F)
Ranks of distances	5	0.4898	0.09796	124.6	**< 2e-16**
Residuals	2994	2.3548	0.00079		

Tukey Honest Significant Difference

	3	5	6	10	12
2	0.99	**0.00**	**0.00**	**0.00**	0.28
3		**0.00**	**0.00**	**0.00**	0.37
5			**0.00**	0.99	**0.00**
6				**0.00**	**0.00**
10					**0.00**

Literature cited

R Development Core Team (2015) R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Rachor E, Nehmer P (2003) Erfassung und Bewertung ökologisch wertvoller Lebensräume in der Nordsee (Survey and assessment of ecologically valuable habitats in the Nordsee). Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany