Vol. 368: 309–310, 2008
doi: 10.3354/meps07728

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published September 25

AS I SEE IT

Misidentification of the Bray-Curtis similarity index

Paul M. Yoshioka*

Department of Marine Sciences, University of Puerto Rico, Mayagüez, PO Box 9013, Mayagüez, Puerto Rico 00681, USA

ABSTRACT: The Bray-Curtis (BC) similarity index is misidentified in various software programs, and the index used by these programs is more appropriately credited to Czekanowski (Cz). The BC and Cz indices differ because abundant species have the greatest impact on the Cz index, while all species (and samples) 'count the same' in the BC index, due to a double standardization of the dataset. To illustrate the difference between these indices, I analyzed the example dataset in the PRIMER manual. In one instance, similarity values differ by 40% (21 vs. 61%). These discrepancies must be considered in comparisons between analyses conducted with software programs that use the true BC index, and programs that use the Cz index but mislabel it as 'BC'.

KEY WORDS: Bray-Curtis index · Czekanowski index · Similarity index · Percent similarity

Marine ecologists often use the Bray-Curtis (BC) index to determine similarities between samples. The formulation for the BC index in various software programs is incorrect and should be attributed to Czekanowski (1909), according to Goodall (1978). The distinction is important because results can differ considerably between the Czekanowski (Cz) and BC indices. Both indices can be classified as measures of 'percent similarity', and for both indices the similarity, S_{jk}, between samples j and k, can be expressed as:

$$s_{jk} = 100\frac{\sum_{i}^{p} 2\min(y_{ij}, y_{ik})}{\sum_{i}^{p}(y_{ij} + y_{ik})}$$

where y_{ij} and y_{ik} represent measures of species i in samples j and k, $\min(y_{ij}, y_{ik})$ is the minimum of y_{ij} and y_{ik}, and p is the number of species. The key difference lies in the manner in which y_i is measured. The Cz index employs quantities (e.g. cover or ind. m^{-2}) for y_i. As a result, the Cz index is more strongly affected by abundant species than less abundant species. In contrast, less abundant species (as well as samples with lower total abundances) have a relatively greater effect on the BC index. This property of the BC index arises from a 'double standardization' procedure, which gives equal weight to all species and samples (Goodall 1978). According to Bray & Curtis (1957), values for species i are first standardized by scoring abundances as a percentage of the maximum value attained by species i over all samples. For the second standardization, the species scores are then rescored as a percentage of the sample totals. Also, as noted by Bray & Curtis (1957), S_{jk} of the BC index can be expressed as $\Sigma_{\min}(y_{ij}, y_{ik})$, a format essentially precluded to the Cz index.

The relative merits of the BC and Cz indices in identifying ecologically meaningful patterns between samples is beyond the scope of this commentary. However, with respect to data characteristics, Whittaker (1967) notes that the BC procedure is especially useful when different categories of measurements are used for different species. In determining similarities between samples, Bray & Curtis (1957) were compelled to standardize species scores in order to combine measures of (1) density, (2) basal area of trees, and (3) frequency of occurrences of shrubs and herbs. In contrast, the Cz index can only be used when abundances of all species are expressed by the same measure (e.g. densities of individuals). Note that the

*Email: p_yoshioka@cima.uprm.edu

Table 1. Abundances of marine benthic species in the example dataset in Table 2.1, p. 2-2 in the PRIMER manual (Clarke & Warwick 2001); S: sample no.

Species	S1	S2	S3	S4
Echinocardium	9	0	0	0
Myriochele	19	0	0	3
Labidoplax	9	37	0	10
Amaeana	0	12	144	9
Capitella	0	128	344	2
Mytilus	0	0	0	0

Table 2. Czekanowski (Cz) and Bray-Curtis (BC) index values (percent similarity) for the dataset in Table 1

Sample	Index	S1	S2	S3
S2	Cz	8		
	BC	11		
S3	Cz	0	42	
	BC	0	31	
S4	Cz	39	21	4
	BC	43	61	14

BC index can also be used when all species are expressed by the same measure.

To illustrate the differing features of the Cz and BC indices, I recalculated the example dataset in the PRIMER manual (Table 2.1 in Clarke & Warwick 2001; Table 1 here). The abundance driven nature of the Cz index is demonstrated by the most abundant taxon, *Capitella*, being responsible for highest similarity (42%) occurring between Samples 2 and 3 (Table 2). In contrast, similarities between Samples 2 and 4 are intermediate with the Cz index (21%), but highest for the true BC index (61%). The latter result is attributable to *Labidoplax*, which is most abundant in Sample 2 and also abundant relative to the other species in Sample 4. This example shows that patterns of percent similarity between samples, and the subsequent interpretation of such patterns, can differ greatly between the Cz and BC indices.

Finally, it must be emphasized that the confusion between the BC and Cz indices extends beyond PRIMER; publications on similarity indices such as Bloom (1981) and Ludwig & Reynolds (1988) have also misidentified the Cz index as the BC index. Such software problems are not unprecedented (Miller 2006). Of greater concern is that this misidentification has largely gone unrecognized for decades, indicating an undiscerning 'cookbook' approach to software programs and ecological indices among marine ecologists as a group. The true BC index is used widely in terrestrial ecology, and the Cz index is used widely in marine ecology; misidentification of the latter as 'BC' can mislead ecologists into making unwarranted comparisons.

Editorial responsibility: Matthias Seaman, Oldendorf/Luhe, Germany

Acknowledgements. This paper is dedicated to the late E. W. Fager for his instruction on the interrelationships between ecology and data analyses. An anonymous reviewer provided very thorough and informative comments on the manuscript.

LITERATURE CITED

Bloom SA (1981) Similarity indices in community studies: potential pitfalls. Mar Ecol Prog Ser 5:125–128

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27: 325–349

Clarke KR, Warwick RM (2001) Change in marine communities: an approach to statistical analyses and interpretation. PRIMER-E, Plymouth

Czekanowski J (1909) Zur differential Diagnose der Neandertalgruppe. Korrespbl dt Ges Anthrop 40:44–47

Goodall DW (1978) Sample similarity and species correlation. In: Whittaker RH (ed) Ordination of plant communities. W Junk, Boston, MA

Ludwig JA, Reynolds JF (1988) Statistical ecology. John Wiley & Sons, New York

Miller G (2006) A scientist's nightmare: software problem leads to five retractions. Science 314:1856–1857

Whittaker RH (1967) Gradient analysis of vegetation. Biol Rev 42:207–264

Submitted: July 31, 2008; Accepted: September 5, 2008
Proofs received from author(s): September 10, 2008