Vol. 372: 307–309, 2008
doi: 10.3354/meps07842

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published December 9

REPLY COMMENT

The Bray-Curtis similarity index remains misidentified: Reply to Somerfield (2008)

Paul M. Yoshioka*

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

ABSTRACT: Somerfield (2008, Mar Ecol Prog Ser 372:303–306) provides numerous reasons for refuting my contention that the Bray-Curtis (BC) similarity index is misidentified by some software programs. He argues that I confuse the pretreatment of data with the calculation of similarity, and ends with the assertion that software programs are not the problem. I argue that the pretreatment of data is 'built into' the BC index, and that software packages are a major source for propagation of the misidentified BC index. Much more importantly, I think that Somerfield's arguments are peripheral to, and do not disprove, the key contention that the index in some software programs differs from the index described by Bray & Curtis (1957). Nevertheless, I agree with Somerfield's sentiment that the term Bray-Curtis should be maintained in community ecology. With respect to indices of percentage similarities I think that Bray and Curtis deserve greater recognition for developing a double standardization procedure that assigns equal weights to species and samples.

KEY WORDS: Bray-Curtis · Percentage similarity · Double standardization · Software errors

Somerfield (2008) provides an extensive list of criticisms against my contention (Yoshioka 2008) that various sources misidentify the Bray-Curtis (BC) index of percentage similarity. I think that Somerfield (2008) overlooks the thrust of my commentary, and before addressing his criticisms I reiterate the core premise upon which my argument rests: namely, that the BC index is misidentified when it differs from the index described by Bray & Curtis (1957).

As stated in Yoshioka (2008, p.309), a general format for indices of percentage similarities is:

$$S_{jk} = 100\frac{\sum_{i}^{p} 2\min(y_{ij}, y_{ik})}{\sum_{i}^{p}(y_{ij} + y_{ik})} \quad (1)$$

where S_{jk} is the similarity between samples j and k, y_{ij} and y_{ik} are 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 problem occurs when sources define the BC index as Eq. (1) used with counts (biomass, frequencies of occurrence, etc.). This rendition of Eq. (1) differs from the 'true' BC index described by Bray & Curtis (1957). For example, following their double standardization procedure, Bray & Curtis' (1957) explicitly state that their index can be simplified to W, which when expressed in the notation of Eq. (1) is:

$$S_{jk} = W = \Sigma\min(y_{ij}, y_{ik}) \quad (2)$$

where y_{ij} and y_{ik} are expressed as percentages of the sample totals, rather than counts. Since Eq. (1) is equivalent to Eq. (2) only if all samples total 100, I argued that Eq. (1) used with counts does not represent the true BC index.

In a larger context, my argument for misidentification follows the convention in science that cited techniques should be rigorously replicated to ensure the reproducibility of results. Thus, citations of the BC index should follow the methodology described by Bray & Curtis (1957). Changes, if any, should be

*Email: p_yoshioka@cima.uprm.edu

acknowledged. For example, Goodall (1978) prefaces his description of the BC index by calling it the 'Wisconsin modification' of Eq. (1).

I demonstrated that the misidentification of the true BC index is of concern by analyzing an example dataset in the PRIMER manual. Similarities differed greatly between the true and misidentified BC indices (Yoshioka 2008, Table 2) because the double standardization procedure in the true BC index assigns equal weights to species and samples in contrast to the abundance-driven results obtained when Eq. (1) is used with counts. Somerfield (2008) does not dispute the cause of the differences, or that results differ greatly between indices. As an aside, I note that the distinction between the true and misidentified BC indices would merit little attention if both approaches produced similar values, as is the case with the Kulczynski index and Eq. (1) used with counts (Clarke & Warwick 2001).

Somerfield (2008) objects to calling Eq. (1) used with counts as the Czekanowski (1909) index. He traces Eq. (1) to Steinhaus, and suggests that Bray & Curtis (1957) erroneously attributed their index to Gleason (1920). I think that Bray & Curtis (1957, p. 329) referred to Gleason (1920) in a different context by restricting comparisons to the Gleason index 'when the sum of score values is relative and equals 100'. In other words, I think that Bray & Curtis (1957) used the Gleason index largely to highlight their standardization procedure.

The discussion of index names diverts attention from a more important issue: confusion is caused by calling different indices by the same name, not by calling the same index by different names. For instance, I was confused by the statement of Pandolfi & Jackson (2001, p. 54) that BC indices 'are dominated by those taxa that attain high abundance'. This situation was clarified when I learned that Pandolfi & Jackson (2001) were referring to the PRIMER version of the BC index.

In regard to index names, Somerfield (2008) also argues that it is incorrect to confuse the pretreatment of data with calculations of similarities. However, the separation of the pretreatment of data and calculations of similarities is an artificial distinction in this particular case because Bray & Curtis (1957) equated their index to Eq. (2), which necessitates that sample totals are standardized to 100%. Somerfield (2008) also states that the pretreatment of data by various down-weighting transformations ultimately leads to abundances being converted to presence/absence data, and calls Eq. (1) used with presence/absence data the Sorenson (Dice) index, thereby contradicting his dictum about the pretreatment of data and index names: 'few ecologists would accept that it is logical for the name of a coefficient to alter depending on the data used to calculate it' (Somerfield 2008 p. 303). By the same reasoning that Somerfield (2008) equates the Sorenson (Dice) to Eq. (1) used with presence/absence data, I feel that Eq. (1) used with counts should be called the Czekanowski (Steinhaus?) index. More importantly, I contend that the BC index should be equated to Eq. (1) only if data is standardized according to Bray & Curtis (1957).

The reference in Somerfield (2008) to correspondence analysis raises a parallel situation with the true and misidentified BC indices. As with the true BC index, double standardization is built into correspondence analysis; and as with Eq. (1) used with counts, the pretreatment of data is not a requisite of principal component analysis. Just as the true BC index can be derived from Eq. (1), correspondence analysis can be derived from principal component analysis by the pretreatment of data (Ludwig & Reynolds 1988). Also, as with the true and misidentified BC indices, results differ greatly between correspondence analysis and principal component analysis. Thus, the contention in Somerfield (2008) that Eq. (1) used with counts should be called the BC index because of data pretreatment, leads to the untenable conclusion that principal component analysis should be called correspondence analysis.

I did not state that use of the BC index is completely distinct between terrestrial and marine ecologists in use. As a marine ecologist, I think of the true BC index when the BC index is cited; conversely, some terrestrial ecologists undoubtedly think of Eq. (1) used with counts as the BC index. Due to the influence of Bray, Curtis, Goodall, Whittaker and their associates, I think that terrestrial plant ecologists are generally more cognizant of the true BC index compared to marine ecologists. However, because overlap occurs between terrestrial and marine ecology, I suspect that confusion exists because — unlike the case of Pandolfi & Jackson (2001) — it may be impossible to determine whether the misidentified or true BC index was used.

I also disagree with the argument in Somerfield (2008) that the misidentification of the BC index is not a problem of software or its use. Without question, software contributes to this problem because users can obtain similarity values while being ignorant of the mathematical procedures involved. Software programs also represent an influential agent for the promulgation of errors, as exemplified by the 'creeping fox terrier clone' syndrome cited by Somerfield (2008). The fact that the misidentified BC index is 'well understood' to be the BC index by the vast majority of ecologists according to Somerfield, only underscores the influence of software programs. As

suggested by Somerfield (2008), errors should be corrected by checking the original source, which in this case is the original description by Bray & Curtis (1957).

Finally, I fully concur with the conclusion in Somerfield (2008) that the use of the term Bray-Curtis should be maintained in community ecology. Bray & Curtis (1957) introduced the polar ordination technique and provided critical evidence supporting the continuum (individualistic) view of communities. In addition, I think that Bray & Curtis (1957) deserve greater recognition for their index, which to my knowledge is unique among indices of percentage similarities in assigning equal weights to species and samples.

Acknowledgements. A NOAA grant for a coral reef ecosystem study (NA170P2919) funded my participation at the recent International Coral Reef Symposium where I became aware of the extent to which the Bray-Curtis index is misidentified by marine ecologists.

Editorial responsibility: Matthias Seaman, Oldendorf/Luhe, Germany

LITERATURE CITED

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

Gleason HA (1920) Some applications of the quadrat method. Torrey Bot Club 47:21–33

Goodall DW (1978) Sample similarity and species correlation. In: Whittaker RH (ed) Ordination of plant communities. W. Junk, Boston, MA, p 101–149

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

Pandolfi JM, Jackson JBC (2001) Community structure of Pleistocene coral reefs of Curacao, Netherlands Antilles. Ecol Monogr 71:49–67

Somerfield PJ (2008) Identification of the Bray-Curtis similarity index: Comment on Yoshioka (2008). Mar Ecol Prog Ser 372:303–306

Yoshioka PM (2008) Misidentification of the Bray-Curtis similarity index. Mar Ecol Prog Ser 368:309–310

Submitted: November 20, 2008; Accepted: November 24, 2008
Proofs received from author(s): November 26, 2008