

Braun-Blanquet data in ANOVA designs: comparisons with percent cover and transformations using simulated data

Bradley T. Furman^{1,*}, Erin H. Leone², Susan S. Bell³, Michael J. Durako⁴,
Margaret O. Hall¹

¹Florida Fish Wildlife Conservation Commission, Florida Fish and Wildlife Research Institute, 100 Eighth Avenue, Southeast, St. Petersburg, FL 33701, USA

²Florida Fish Wildlife Conservation Commission, Florida Fish and Wildlife Research Institute, Center for Biostatistics and Modeling, 1105 SW Williston Road, Gainesville, FL 32601, USA

³University of South Florida, Department of Integrative Biology, 4202 East Fowler Avenue, Tampa, FL 33620, USA

⁴University of North Carolina at Wilmington, Center for Marine Research, 7205 Wrightsville Avenue, Wilmington, NC 28403, USA

ABSTRACT: The Braun-Blanquet (BB) cover-abundance scale is used to visually estimate community composition and species dominance. An 8-division variant was developed for benthic systems in the 1990s; the capacity to speed collection of seagrass coverage data led to its adoption by several large-scale monitoring programs in the USA. However, debate regarding how best to treat ordinal BB data in statistical analysis has stymied progress in the comparison of status and trends. Methods specific to ordinal data exist; however, they have generally been ignored in favor of transformation to percent cover or the use of BB categories as continuous data in parametric statistics and multivariate ordination. To quantify behavior of BB data in 1-way ANOVA, we conducted a series of data simulations using percent cover, BB scores and 3 metric-scale transformations as competing dependent variables in iterated 2-group contrasts. Simulations followed the design of the Fisheries Habitat Assessment Program (FHAP) and covered full ranges of within- and between-group variation. We empirically estimated Type I error and proportional deviance in effect size as measures of performance. Finally, we compared 6 yr of FHAP data to the simulations to identify scenarios likely to be encountered by seagrass ecologists. BB scores performed well as a proxy for continuous data and log-linear transformation allowed more precise effect size estimation. Our results highlight the need for high levels of replication in benthic sampling and provide empirical evidence for the statistical reliability of BB data in parametric analysis.

KEY WORDS: Braun-Blanquet · ANOVA · Categorical · Data simulation · Statistical power · Data transformation · Benthic cover · Seagrass · Benthic monitoring · Percent cover

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

The Braun-Blanquet (BB) cover-abundance scale is a 5- to 9-division ordinal alphanumeric scale commonly used in vegetation science to visually estimate community composition and species dominance (Braun-Blanquet & Pavillard 1922, Braun-Blanquet 1932). First developed by Braun-Blanquet

in the 1920s, the technique evolved from earlier use of areal percentages to define plant community types or 'syntaxa', and quickly became the modern standard for recording relevé data (Braun-Blanquet 1932, Mueller-Dombois & Ellenberg 1974, Adamczyk et al. 2013). The BB scale and its accepted variants revolutionized the nascent field of phytosociology and set the stage for decades of

globally accumulated vegetation data (Leps & Hadincova 1992).

As a field assay, the appeal of the BB scale was obvious: it provided fast and efficient means of non-destructively sampling large plots (Braun-Blanquet advocated up to 10 m²) for both species abundance and cover, allowing investigators to sample quickly in multistory environments and to integrate information over larger spatial scales than would have been practical using quantitative metrics, such as above- and below-ground biomass, productivity or leaf-area index (Braun-Blanquet 1932, Dobert et al. 2015, Herpigny & Gosselin 2015). That BB scores co-varied with at least some of these measures added to the scale's utility and widespread popularity (Dobert et al. 2015). Moreover, by formalizing the abundance-cover coefficient at intuitive breaks (i.e. sparse, plentiful, 5, 25, 50, 75%), the scale reduced observer subjectivity and minimized inter-experimenter variation (Edwards & Tinker 2009).

The central role of the BB scale in phytosociology led to a lasting presence in terrestrial plant ecology; however, it was rarely used in marine systems until a modified version was introduced to seagrass ecology by Kenworthy et al. (1993; but see Watling et al. 1978, Niell 1979). The 8-division scale preserved the abundance-to-cover sequence put forth by Braun-Blanquet (1932), but replaced alphanumeric category labels, 'r' and '+', with fully numeric ones, 0.1 and 0.5, respectively. Generally applied using smaller, 0.25 to 1 m² quadrats, the marine variant of the BB score allowed investigators to census far more sites, with greater replication over larger geographic distances, than could have been achieved using alternative methods such as seagrass shoot counting, benthic coring or point-intercept coverage estimation (Kenworthy et al. 1993, Rose et al. 1999, Fourqurean et al. 2001, Armitage et al. 2006, Hall et al. 2007). For these reasons, the modified BB scale was chosen to assess benthic macrophyte distribution and abundance by 2 of the largest and longest running seagrass monitoring programs in the USA: the Fisheries Habitat Assessment Program (FHAP; 1995 to the present) in Florida Bay and the Florida Keys National Marine Sanctuary Seagrass Monitoring Program (FKNMS SMP; 1995 to the present) in the Florida Keys. Both programs currently represent 20+ yr repositories of benthic coverage data recorded on the BB scale (Fourqurean et al. 2001, Durako et al. 2002).

Despite the BB scale's undeniable value to vegetation and seagrass ecology, and the ongoing need to integrate contemporary sampling with historical datasets, the use of BB data becomes problematic

during statistical analysis. BB data are (1) zero-inflated—owing to patchy distributions of most target species (Herpigny & Gosselin 2015), (2) non-numerical—'r' and '+' have no obvious numerical equivalent (Podani 2006) and (3) categorical—although conversion to an ordinal scale can be made (van der Maarel 1979). As Podani (2006) pointed out, the BB scale arrived well before computer-aided statistics or modern sampling theory, so the implications of semi-quantitative data did not factor into its adoption. In the marine realm, many of the earliest uses of the BB score were either strictly qualitative (Niell 1979) or relied on derived metrics that were themselves continuous (but see Watling et al. 1978), thereby circumventing categorical issues altogether (Kenworthy et al. 1993, Martin-Smith 1993).

To address limitations in the quantitative analysis of BB scale data, 5 solutions have been proposed: (1) convert to presence/absence (van der Maarel 1979), (2) derive and analyze continuous metrics, such as frequency, abundance and density (Kenworthy et al. 1993), (3) convert to ordinal scale, then only employ methods appropriate to ordered data (Podani 2005, Ricotta & Feoli 2013), (4) convert from ordinal to continuous and apply conventional statistics (van der Maarel 2007), including those tailored to zero-inflated data (Herpigny & Gosselin 2015), and for the marine variant, (5) ignore data scale altogether and treat raw BB scores as though continuous (Rose et al. 1999, Armitage & Fourqurean 2006). All of these approaches have demonstrated success in describing patterns in plant community data. However, an ongoing debate remains regarding the proper balance between mathematical admissibility (Podani 2006, Ricotta & Feoli 2013), which precludes the use of ratio-scale methods on ordered data, and the value of reinforcing researcher intuition, since the use of such methods tends to recover anticipated patterns (Clymo 1980, Ricotta & Avena 2006, but see Podani 2007).

Podani (1984) suggested that the acquisition of field data can be quite imprecise and may in fact be the limiting step in any analysis of BB data. Here, we examine how much information is lost with respect to quantifying species cover when a BB scale is used. We then assess which of the approaches common in the marine literature performs best when used in simple 1-way ANOVA. Finally, we determine if either of the log-linear transformations proposed by van der Maarel (van der Maarel 2005, 2007) offer advantages over less severe methods. To do this, we used the FHAP sampling design to simulate coverage data and BB scores for 2 sample groups along 3 axes of variation: effect size (difference between group

means), variation about the mean and the midpoint between the means of 2 sample groups. We then compared the simulated data to 6 yr of FHAP seagrass data, and assessed Type I error (α) and proportional deviance in raw mean difference effect size after iterative 1-way ANOVAs between sample populations. This study represents an attempt to answer the question of how best to treat BB scale data in parametric analysis, offering detailed information on the capacity of raw BB data and common transformations to resolve treatment differences (i.e. statistical power) in an experimental design ubiquitous in the marine literature (Underwood 1997).

METHODS

Ecological data

The FHAP program began in 1995 in response to the first large-scale die-off reported for *Thalassia testudinum* in Florida Bay, Florida (Robblee et al. 1991, but see Hall et al. 2016 for a recent recurrence). To track seagrass recovery in discrete sections of the bay, known as ‘basins’ or ‘lakes’ due to their geomorphology and hydrologic isolation by shallow subtidal carbonate mudbanks, bi-annual (1995 to 2004), then annual (2005 to present), monitoring of benthic coverage was conducted in 10 to 17 basins (Fig. 1) using the modified BB scale of Kenworthy et al. (1993; our Table 1). Each basin represented a range of physico-chemical conditions and gradients in Florida Bay, and was partitioned into 29 to 31 tessellated, hexagonal grid cells using algorithms developed by the US Environmental Protection Agency’s (EPA) Environmental Monitoring and Assessment Program (EMAP). Sampling sites, 1 per grid cell, were randomly chosen each year and field-located using handheld GPS. At each site, benthic cover and abundance of all macrophytes—by species, taxonomic group and total—were visually estimated within each of eight 0.25 m² quadrats, haphazardly distributed within a 10 m radius of the anchored vessel. Analyses of the BB and transformed data treated site-level means ($n = 8$) as replicate samples for basin-level comparison ($n = 29$ to 31).

The portion of the dataset used for this study covered the years 2010 through 2015, included 6 spring and 1 fall samplings, was comprised of 3222 sampling

Table 1. Braun-Blanquet scales and common conversions, adapted from van der Maarel (2007). % Cover ranges: the ranges used to convert continuous scale benthic coverage to the marine variant of the Braun-Blanquet scale; BB: the marine variant of the Braun-Blanquet score developed by Kenworthy et al. (1993); BB2: the extended Braun-Blanquet scale from which the ordinal transformations of van der Maarel (1979) originated; TxEl: the updated midpoint conversion first suggested by Tüxen & Ellenberg (1937); OTV: the ordinal transform scale of van der Maarel (1979); vdm05: the logarithmic transformation of van der Maarel (2005); vdm07: the updated logarithmic transformation of van der Maarel (2007)

% Cover ranges	BB	BB2	TxEI	OTV	vdm05	vdm07
[0]	0	0	0	0	0.24	0.23
(0, 0.04]	0.1	r	0.02	1	0.493	0.485
(0.04, 0.2]	0.5	+	0.1	2	1.000	1.00
(0.2, 5]	1	1	2.5	3	2.027	2.064
–	–	2m	–	4	4.110	4.260
(5, 25]	2	2a	15	5	8.332	8.793
–	–	2b	–	6	16.427	18.148
(25, 50]	3	3	37.5	7	34.246	37.457
(50, 75]	4	4	62.5	8	69.427	77.310
(75, 100]	5	5	87.5	9	140.752	159.567

sites and represented data from 25 776 quadrats (8 per site; Fig. 1). For simplicity, only the 6 seagrass species (*T. testudinum*, *Syringodium filiforme*, *Halodule wrightii*, *Ruppia maritima*, *Halophila engelmannii* and *Halophila decipiens*) and the composite category: ‘Total Seagrass’ were used for analysis. These data represent a 6 yr period of post-recovery seagrass community dynamics along with a limited

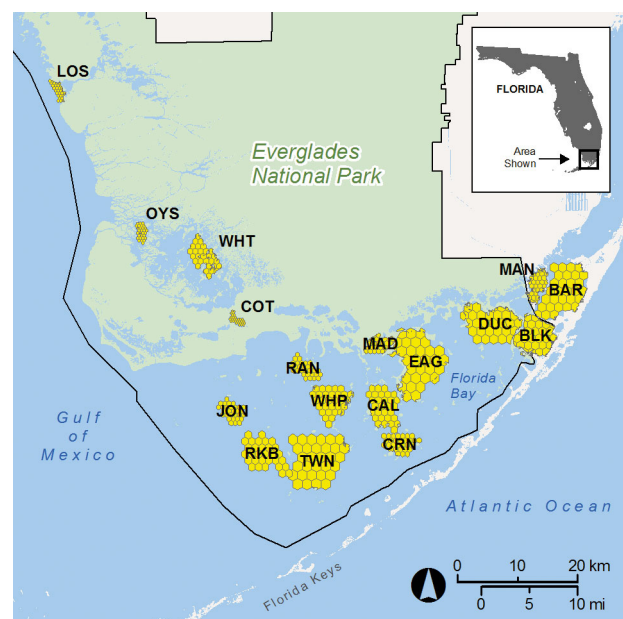


Fig. 1. Fisheries Habitat Assessment Program (FHAP) sampling hexagons included in the simulation study (2010 to 2015)

5-basin fall survey of active die-off during the recent 2015 mortality event (Hall et al. 2016). Therefore, the comparison dataset captured a wide range of naturally occurring variation in seagrass species composition, patchiness and coverage.

Simulated data and transformations

To capture the behavior of BB data where actual percent cover was known, we simulated FHAP methodology for 2 basins using 30 replicate sites per basin and 8 subsample quadrats per site (Fig. 2A). Three statistics were varied: (1) standard deviation, (2) effect size or distance between basin means and (3) midpoint between the means of both sample groups, in a fully orthogonal mixture distributed across a wide range of values (Fig. 2B). Basin means extended from 1 to 100% cover in 1% increments (Pop1: 1 to 99%; Pop2: 3 to 100%) with corresponding influence on effect size (1 to 99%) and midpoint (2 to 99.5%). Standard deviations varied from 1 to 100% in 1% increments from 1 to 5%, and in 5% increments thereafter. Treated as axes in 3-dimensional space, the resulting triangular prism formed a 'performance space' within which all simulated benthic coverage and transformed metrics were later evaluated (Fig. 3A).

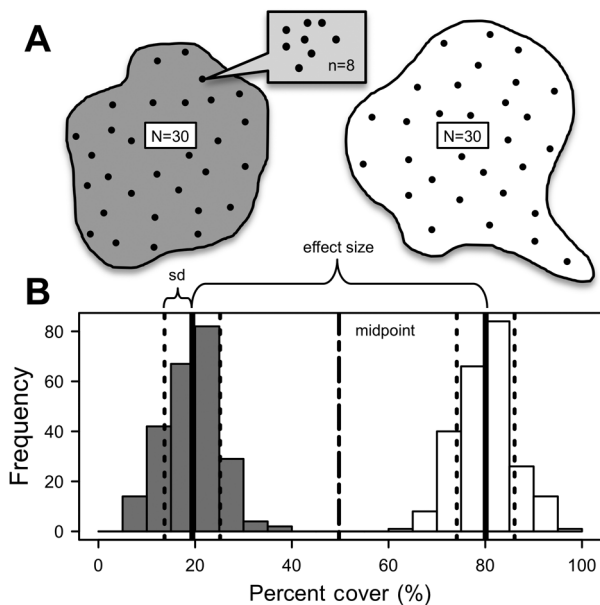


Fig. 2. A graphical overview of the simulation showing (A) the pseudoreplicated ($n = 8$ per site) and replicated ($N = 30$ per basin) sampling of 2 basins following Fisheries Habitat Assessment Program (FHAP) methodology and (B) the 3 underlying population parameters varied during the study: standard deviation (SD), effect size and midpoint between the means of both basins

For each unique combination of standard deviation, effect size and midpoint ($n = 60\,000$), 1000 iterations of FHAP sampling were simulated ($N = 6 \times 10^7$). Quadrat-level percent coverage values were randomly drawn from normal distributions meeting each iterative set of specified parameters. Values outside of $[0, 100]$ were substituted for 0 and 100, respectively. All simulated percent cover data were converted to the BB scale of Kenworthy et al. (1993) using the following ranges: $[0] = 0$, $(0, 0.04] = 0.1$, $(0.04, 0.2] = 0.5$, $(0.2, 5] = 1$, $(5, 25] = 2$, $(25, 50] = 3$, $(50, 75] = 4$ and $(75, 100] = 5$ (Tüxen & Ellenberg 1937, van der Maarel 2007). From raw BB scores, 3 common percent cover transformations were attempted: the median transformation adapted from

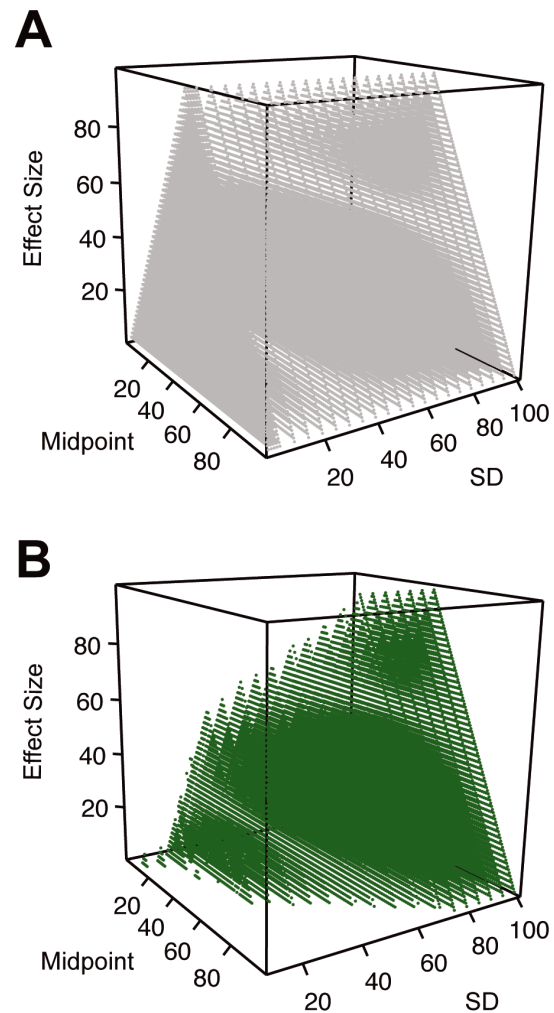


Fig. 3. (A) The 3-dimensional 'performance space' (gray) defined by the systematic sampling of effect size, midpoint and standard deviation (SD) between the simulated sampling of benthic cover between 2 basins and (B) the approximate portion of that space (green) where Fisheries Habitat Assessment Program (FHAP) data were likely to occur

Tüxen & Ellenberg (1937), 'TxEl', and 2 logarithmic transformations offered by van der Maarel (2005, 2007), 'vdm05' and 'vdm07'. Both van der Maarel methods required initial conversion to the ordinal scale following van der Maarel (1979; our Table 1). Formulas for vdm05 and vdm07 were as follows:

$$\ln(C) = \frac{(OTV - 2)}{a} \quad (1)$$

where OTV is the ordinal transform value, C is the approximate percent cover and a is a weighting factor equal to 1.415 and 1.380 for vdm05 and vdm07, respectively (van der Maarel 2007).

For each simulated FHAP scenario, the percent cover, BB score and all 3 transformations (TxEl, vdm05 and vdm07) were successively used as the dependent variable in 1-way ANOVA, testing the main effect of 'basin' on cover, for which the expected outcome in all cases was a significant difference in basin means. To quantify performance of each transformation, we recorded 2 measures of accuracy from each set of iterated scenarios ($n = 1000$ per set): (1) the proportion of significant ANOVAs (at $\alpha = 0.05$), providing an empirical estimate of α and (2) the mean estimated effect size (i.e. the absolute difference between simulated basin means), which provided an integrated measure of precision for location statistics from each basin. For further comparison, mean effect size differences were converted to proportion deviance values following:

$$d = \left| \frac{(ES_{\text{simulated}} - ES_{\text{known}})}{ES_{\text{known}}} \right| \quad (2)$$

where ES is the raw mean difference effect size estimated from each simulated scenario and the known ES for that scenario. Because BB scores are categorical data but were treated as continuous for the analyses, the actual effect size was converted to the BB scale (Kenworthy et al. 1993) prior to calculating deviance.

Mapping ecological data to the performance space

Because the performance space was based on population-level parameters and because we lacked true percent cover for the FHAP data, our capacity to map FHAP data directly to the performance space was limited. However, to best approximate ecological conditions we matched FHAP BB score behavior at the site and basin scales to corresponding parameter combinations from the simulation. To do this, we first simulated FHAP-style sampling of single basins, pre-

served the ranges in site-level mean and standard deviation for each unique combination ($n = 1000$ per combination; $n = 2400$ combinations; $N = 2.4 \times 10^7$) and matched those simulated ranges to observed ranges from FHAP data. Through this exercise we were able to map the standard deviation from the simulated performance space to FHAP data. To calculate midpoint and effect size locations, we identified simulation scenarios where both simulated basins were within FHAP data ranges. The resulting dataset functioned as a liberal estimate of ecological space (hereafter, 'FHAP zone') within the simulated performance space, thus allowing an honest assessment of BB score and transformation behavior likely to be encountered by investigators using field data.

Statistical analysis

All data simulations and univariate statistical analyses were conducted using the statistical software, R version 2.14.1 (R Development Core Team 2012). One- and 2-way ANOVAs were conducted using the *avov* function from the package 'stats'. Post hoc comparisons were made using Tukey's honest significant difference test as implemented by the package 'agricolae'. In all cases, statistical significance was assessed at $\alpha = 0.05$.

RESULTS

The FHAP zone

Range matching between sample statistics (\bar{x} and s) derived from FHAP data and population parameters (μ and σ) using simulation-derived statistics yielded a reasonable approximation of the FHAP zone, or the region of the performance space wherein actual ecological data were likely to occur (Fig. 3B). The FHAP zone occupied 51% of the performance space (30 732 of 60 000 sets of simulations) and was shifted somewhat toward the higher end of the standard deviation and lower end of the midpoint axes across a wide range of modeled effect sizes. Mean (± 1 SD) 'Effect Size', 'Midpoint' and 'SD' occupied by FHAP data were $31.3 \pm 23.6\%$, $46.4 \pm 20.4\%$ and $69.1 \pm 32.3\%$, respectively. As expected for seagrass coverage data, there was a positive relationship between midpoint distance (i.e. the magnitude of the true mean percent seagrass cover between 2 basins) and standard deviation along the upper range of midpoints for any set of standard deviation. In other

words, contrasts between basins of increasing sea-grass coverage tended to involve greater degrees of within-basin variance, perhaps owing to patchiness at the highest end of the coverage spectrum. Also, due to natural variance among the 3 performance space axes, minimum SD (10%) was 5 to 10× that of Midpoint (2%) or Effect Size (1%), despite all 3 covering the balance of their modeled ranges at 10 to 100%, 2 to 99.5% and 1 to 99%, respectively.

Simulated ANOVA

The 4 transformations examined for performance in simple 1-way ANOVA exhibited efficacy comparable to that of actual percent cover. For the simulated data, 73.5% of tests found significant differences between basins using percent cover. Once converted to BB scores, 70.5% of tests were significant, while 70.4, 70.0 and 70.0% were significant using the TxEl, vdm05 and vdm07 transformations, respectively. When only considering simulations that fell within the range of data observed under FHAP, the likelihood of detecting significant differences fell closer to 50%, where higher variance may have begun to interact with limitations of the simulated FHAP sampling; i.e. statistical power may have been adversely affected by sample density as well as metric or transformation idiosyncrasy. For example, over 26.5% of the performance space sample replication was insufficient to resolve basin differences using 1-way ANOVA. This rose to 45% within the FHAP zone. Yet, there were very few instances where only percent cover was needed (2.9 and 0.8% for performance space and FHAP zone, respectively), and these tended to be aggregated along breakpoints in the BB scale at the lowest ends of the SD and Effect Size axes (Fig. 4).

When considering BB transformations along a continuum of increasing severity (BB < TxEl < vdm05 < vdm07), raw BB score, treated as though a continuous variable, outperformed the 3 transformations across both the performance space at 70.5% and within the FHAP zone at 54% of simulated scenarios. For cases where $\alpha = 0.05$ was achievable, BB score was the transformation of choice for 99.6% of the FHAP zone and 99.8% overall (Fig. 4A). Only along a single plane between where sampling was insufficient to resolve basin differences (i.e. where even percent cover was insufficient) and where BB data performed well were alternate transformations needed, and here only TxEl and vdm05 were occasionally found (Fig. 4B).

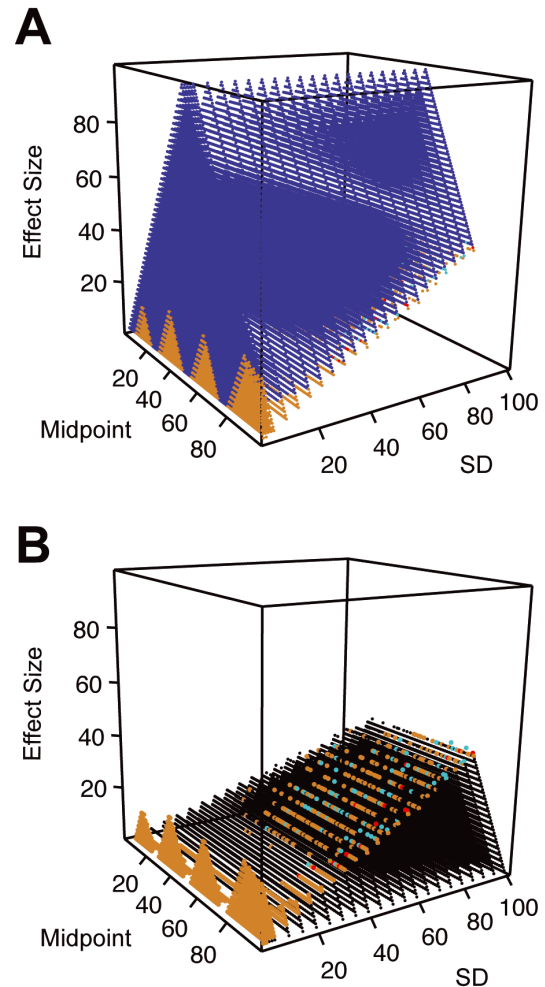


Fig. 4. Results of the ANOVA simulation showing (A) the identity of the least severe transformation (BB [dark blue] < TxEl [turquoise] < vdm05 [red] < vdm07 [none]) needed to resolve basin differences at an empirical $\alpha = 0.05$, and (B) the portion of the performance space where actual percent cover (light brown) was needed and where percent cover (black) was incapable of effective ANOVA. For clarity, the TxEl and vdm05 transformations, as well as percent-cover-only locations, are shown in both panels

Estimation of effect size

The capacity of transformed percent cover data to estimate effect size (i.e. a proxy location statistic) was captured as the proportional deviance between mean effect size estimated during simulation and the known effect size for each simulation scenario. Patterns among transformations were different within, as opposed to outside, the FHAP zone; this interaction between zone and transformation was significant in a 2-way ANOVA ($F_{12,4633} = 1723.8$, $p < 0.001$; Fig. 5). For percent cover, deviance was higher within the FHAP zone (0.445 ± 0.148 , SD) than out-

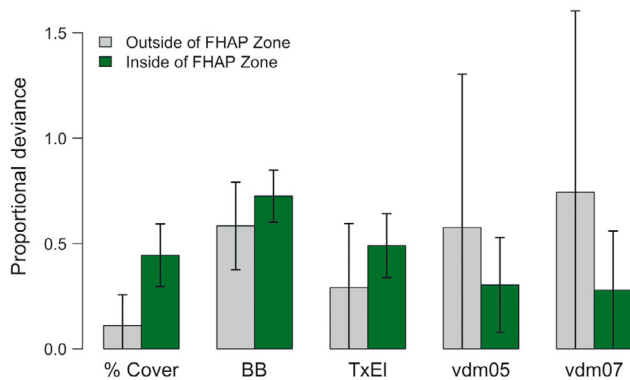


Fig. 5. Interaction plot of proportional deviance between simulated and known raw mean difference effect sizes between 2 simulated FHAP basins. Mean \pm 1 SD deviance for actual percent cover and 4 transformations are shown as calculated within (dark green) and outside (gray) the FHAP zone

side (0.111 ± 0.146), as it was for BB score (FHAP zone: 0.725 ± 0.124 ; outside FHAP: 0.584 ± 0.208) and TxEl (FHAP zone: 0.490 ± 0.152 ; outside FHAP: 0.291 ± 0.304). However, this pattern, consistent with higher variance and ANOVA limitations within the FHAP zone, reversed for the van der Maarel transformations, which both exhibited lower proportional deviance within the FHAP zone. A 1-way ANOVA by Transformation Type within the FHAP zone ($F_{4,153655} = 25912$, $p < 0.001$) and post hoc comparison tests revealed all 4 transformations to be significantly different from each other and raw percent cover, with vdm07 outperforming all others with a proportional deviance of 0.279 ± 0.280 (61% lower than raw BB score and 43% lower than TxEl).

In the performance space, the distribution of transformations, prioritized using the minimum proportional deviance, exhibited clear dominance of vdm07 at 46% of modeled scenarios (Fig. 6A), rising to 75% within the FHAP zone (Fig. 6B). Conversely, raw BB was the optimal transformation at only 2% of the performance space and only 6 of 30 732 observations within the FHAP zone. TxEl and vdm05 occupied 33 and 19% of the performance space, and 9 and 16% of the FHAP zone, respectively.

DISCUSSION

For nearly a century, the Braun-Blanquet cover-abundance scale and its accepted variants have offered practical solutions for recording plant community composition and coverage, yet as semi-quantitative data, their use in modern statistical tests remains problematic (Podani 2006). We used a series

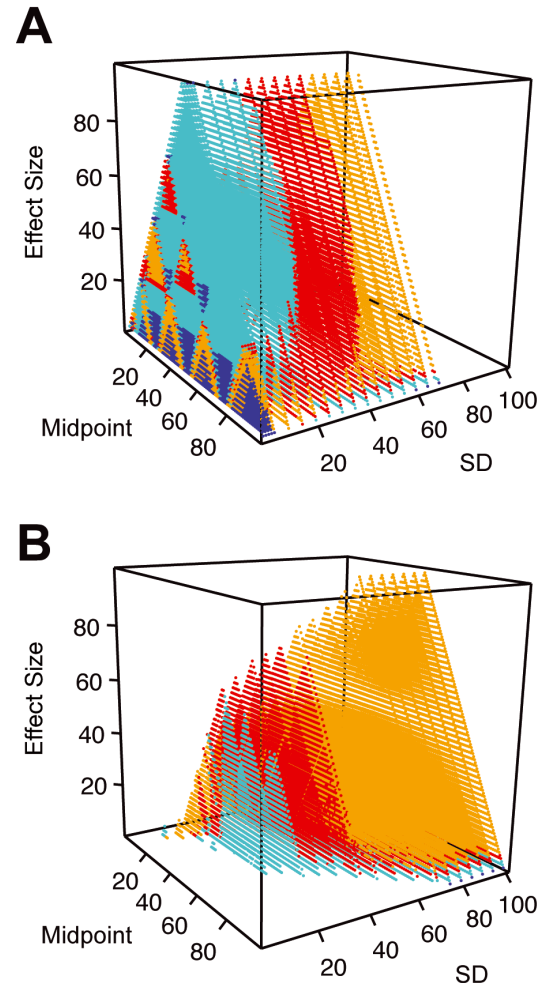


Fig. 6. The distribution of prioritized transformations (BB [dark blue], TxEl [turquoise], vdm05 [red], vdm07 [orange]) both (A) outside and (B) within the FHAP zone following data simulation. Priority was given to the transformation that had the least proportional deviance in raw mean difference effect size

of 2-group data simulations to empirically estimate F -statistic performance in 1-way ANOVA using percent cover, BB scores and 3 BB transformations (TxEl, vdm05, vdm07) as competing dependent variables. Modeled scenarios spanned more or less complete ranges of within-group population parameters (mean and SD) and among-group characteristics (midpoint and effect size) providing a holistic picture of transformation effectiveness. By mapping field-acquired benthic macrophyte data to the simulated performance space we were able to further refine our assessment to situations encountered by seagrass ecologists working in sub-tropical systems (i.e. FHAP zone). Despite improper data type, the 8-division marine variant of the BB scale (Kenworthy et al. 1993) performed well in parametric statistical tests

when treated as continuous data without prior transformation. Surprisingly, the proportion of total space where BB data were effective (i.e. $\alpha \leq 0.05$) was only 3% less than that of true percent cover, and was actually better than any of the 3 metric transformations. Within the FHAP zone, the difference fell to only 0.9% and there was a similar spread among competing transformations. This demonstrates that for the preponderance of typical monitoring program analyses, data collected in the BB scale can be directly analyzed using ANOVA without significant penalty to statistical power.

Although BB data successfully capture important aspects of sample dispersion, they unfortunately provide nonsensical estimates of sample group location. For example, a group mean of 2.5 has neither appropriate units nor real world analogy (at least in terms of percent cover), and so graphical representations of the data and group differences (i.e. raw mean difference effect sizes) are difficult to interpret directly. The simplest remedy is to remap BB data back to continuous scales based on percent cover, which is appropriate since areal coverage underpins most of the logic supporting BB data. We tested 3 such transformations and found that the most recent log-linear formula by van der Maarel (2007; vdm07) yielded optimal results, especially within the FHAP zone, where it was favored in 75% of cases. The reasons for this are not clear, but some have argued that conversions beyond 100% might better represent the multi-story nature of vegetation canopies (van der Maarel 2005), something that would certainly apply to the mixed-meadow seagrass and macroalgal communities modeled here. Further, for sample groups on the upper end of the scale, we speculate that having subsamples greater than 100% might contribute to more accurate replicate means and, by consequence, more accurate sample group means, although we did not test this specifically. Nevertheless, our results are congruent with the idea that log-linear transformations work well in multivariate space (Ricotta & Feoli 2013), because more precise location statistics likely produce better dissimilarity matrices. Because the vdm07 transformation maps BB data onto a continuous metric scale recognizable to ecologists who may not be familiar with BB categories, it has added benefit when graphically representing sample groups in standard plots or when considering quantitative predictions of coverage change (i.e. a 45.6% change in a species coverage is far more interpretable than a 1.2 shift in BB score).

Perhaps our most salient finding was the clear limitation that natural variability places on an otherwise

robust sampling design. Even though the proportion of the FHAP zone that required data in true percent cover for $\alpha = 0.05$ was vanishingly small (0.8%), nearly half (44.9%) of the zone exhibited within-group variability beyond the statistical power of our sampling protocol. This occurred despite percent cover being precisely known for each subsample, something not attainable in the field. Not all of the failed contrasts would have been regarded as ecologically relevant, but it does suggest that the primary challenge in estimating benthic macrophyte community characteristics remains variability among samples rather than the accuracy of individual samples.

This has been appreciated for some time. Dethier et al. (1993) showed that visual estimation techniques allowed for larger sample sizes, reduced overall sampling error and produced more accurate representations of benthic communities, than did those based on more laborious point-count estimates. In addition, random point sampling is more prone to missing rare or isolated species than visual surveys (Dethier et al. 1993). Corroborating this, Edwards & Tinker (2009) compared targeted (a hierarchical suite of approaches using transects and quadrats tailored to multiple taxonomic groups) and coefficient sampling (TxEl-transformed BB quadrat data) and found that the increased replication made possible using visually based BB data overcame issues of sample accuracy. Thus, increased replication reduces sampling error (Dethier et al. 1993) and increases precision (Mellors 1991) more so than comparable efforts using more quantitative (line- or point-intercept sampling) or time consuming (shoot-count) and destructive (core harvesting) methods. Because BB techniques directly mitigate the biggest issue in ecological sampling: achieving sample densities sufficient to resolve group differences, its application in the field provides investigators with a statistical hedge against spatio-temporal variability.

Edwards & Tinker (2009) cautioned against the use of BB data for experiments involving fine spatial grain or small effect sizes. Such concerns have caused trepidation among practitioners weighing a decision to record field data on the BB scale or attempting to analyze BB datasets. It is reasonable to expect deficiencies given the subjective, categorical nature of the scale, and possible human errors associated with decision-making near category transitions. Consistent with this, we observed errors for ANOVA scenarios conducted in the vicinity of BB category breakpoints, at effect sizes less than 20% (mean \pm 1 SD of $9.4 \pm 8.1\%$) or at very low SD (median = 3; 13.8 ± 23.8 ; Fig. 4); however, most of the

FHAP data were found outside of these bounds. Therefore, our simulation provides empirical evidence that these common concerns may not be appropriate for real-world applications of BB techniques in benthic macrophyte systems, especially when subsamples are taken. Overall, the enhanced replication outweighs the risk.

Debate regarding mathematical admissibility of BB data in parametric univariate statistics and multivariate ordinations (Ricotta & Avena 2006, Podani 2007) is far from resolved, and while the current study does not directly address it, we can comment on how BB data should be regarded by field ecologists and coastal managers tasked with collecting long-term benthic coverage data at landscape scales. First, because the use of BB data in modern analysis is not supported by statistical theory, their behavior must be empirically estimated. By doing so, we can gain confidence in how the metric (when used as such) will perform in a given sampling design. Our results suggest that BB data can be used as a continuous metric in simple ANOVA and increased accuracy is attainable using log-linear (vdm07) transformation. Second, given ever-present sampling limitations and pervasive underlying natural variation, we argue that the value of BB data may be less a matter of recovering patterns perceived *a priori* by an investigator (Clymo 1980, Ricotta & Avena 2006, Ricotta & Feoli 2013) than it is about statistical reliability, and the consistency with which a given test can resolve true sample group differences, given ever-present sampling limitations and pervasive underlying natural variation.

This study builds upon the idea that sampling speed (leading to greater replication) can be more important than sample accuracy (Mellors 1991, Dethier et al. 1993, Edwards & Tinker 2009), showing that visually based semi-quantitative BB techniques are an effective tool in mitigating variability in seagrass systems. BB data should therefore not be viewed as a simple alternative scale to, say, percent cover, but as a means of achieving greater sample densities. When used as such, we have shown that the resulting data can function effectively in common parametric statistics.

Acknowledgements. The authors thank the South Florida Water Management District for funding under the Restoration, Coordination and Verification Program, and the Everglades National Park for substantial logistical support of the FHAP monitoring program. FHAP would not have been possible without the assistance of many UNCW graduate students and FWC seagrass team members over the past 20+ years, particularly Manuel Merello, Donna O'Dell, Katie

Toth, Mike Wheeler, Robin Jung, Nathan Gavin, Brooke Landry and Jennifer Kunzelman Hart. Data simulations were run with computational support from the University of Virginia, provided by Dr. Eric Bricker. We also thank Drs. Amber Stubler, Kara Radabaugh, Colin Shea and 2 anonymous reviewers for their insightful comments and assistance on earlier drafts of this work.

LITERATURE CITED

- ✦ Adamczyk JJ, Kurzac M, Park YS, Kruk A (2013) Application of a Kohonen's self-organizing map for evaluation of long-term changes in forest vegetation. *J Veg Sci* 24: 405–414
- ✦ Armitage AR, Fourqurean JW (2006) The short-term influence of herbivory near patch reefs varies between seagrass species. *J Exp Mar Biol Ecol* 339:65–74
- ✦ Armitage AR, Frankovich TA, Fourqurean JW (2006) Variable responses within epiphytic and benthic microalgal communities to nutrient enrichment. *Hydrobiologia* 569: 423–435
- Braun-Blanquet J (1932) *Plant sociology: the study of plant communities*. Authorized English translation of *Pflanzensoziologie*. Translated, revised and edited by GD Fuller and HS Conard. McGraw-Hill, New York, NY
- Braun-Blanquet J, Pavillard J (1922) *Vocabulaire de sociologie végétale*. Romégou et Dehan, Montpellier
- ✦ Clymo RS (1980) Preliminary survey of the peat-bog Hummell Knowe moss using various numerical methods. *Vegetatio* 42:129–148
- ✦ Dethier MN, Graham ES, Cohen S, Tear LM (1993) Visual versus random-point percent cover estimations: 'objective' is not always better. *Mar Ecol Prog Ser* 96:93–100
- ✦ Dobert TF, Webber BL, Sugau JB, Dickinson KJM, Didham RK (2015) Can leaf area index and biomass be estimated from Braun-Blanquet cover scores in tropical forests? *J Veg Sci* 26:1043–1053
- Durako MJ, Hall MO, Merello M (2002) Patterns of change in the seagrass-dominated Florida Bay hydroscape. In: Porter JW, Porter KG (eds) *Linkages between ecosystems in the south Florida hydroscape: the river of grass continues*. CRC, Boca Raton, FL
- Edwards MS, Tinker MT (2009) Monitoring benthic algal communities: a comparison of targeted and coefficient sampling methods. *Algae* 24:111–120
- ✦ Fourqurean JW, Willsie A, Rose CD, Rutten LM (2001) Spatial and temporal pattern in seagrass community composition and productivity in south Florida. *Mar Biol* 138: 341–354
- Hall MO, Madley K, Durako MJ, Zieman JC, Robblee MB (2007) Florida Bay. In: Handley L, Altsman D, DeMay R (eds) *Seagrass status and trends in the Northern Gulf of Mexico: 1940–2002*. US Geological Survey Scientific Investigations Report 2006–5287, p 243–254
- ✦ Hall MO, Furman BT, Merello M, Durako MJ (2016) Recurrence of *Thalassia testudinum* seagrass die-off in Florida Bay, USA: initial observations. *Mar Ecol Prog Ser* 560: 243–249
- ✦ Herpigny B, Gosselin F (2015) Analyzing plant cover class data quantitatively: customized zero-inflated cumulative beta distributions show promising results. *Ecol Inform* 26:18–26
- ✦ Kenworthy WJ, Durako MJ, Fatemy SMR, Valavi H, Thayer GW (1993) Ecology of seagrasses in northeastern Saudi

- Arabia one year after the Gulf War oil spill. *Mar Pollut Bull* 27:213–222
- Leps J, Hadincova V (1992) How reliable are our vegetation analyses? *J Veg Sci* 3:119–124
- Martin-Smith KM (1993) Abundance of mobile epifauna: the role of habitat complexity and predation by fishes. *J Exp Mar Biol Ecol* 174:243–260
- Mellors JE (1991) An evaluation of a rapid visual technique for estimating seagrass biomass. *Aquat Bot* 42:67–104
- Mueller-Dombois D, Ellenberg H (1974) *Aims and methods of vegetation ecology*. John Wiley and Sons, New York, NY
- Niell FX (1979) Structure and succession in rocky algal communities of a temperate inter-tidal system. *J Exp Mar Biol Ecol* 36:185–200
- Podani J (1984) Spatial processes in the analysis of vegetation: theory and review. *Acta Bot Hung* 30:75–118
- Podani J (2005) Multivariate exploratory analysis of ordinal data in ecology: pitfalls, problems and solutions. *J Veg Sci* 16:497–510
- Podani J (2006) Braun-Blanquet's legacy and data analysis in vegetation science. *J Veg Sci* 17:113–117
- Podani J (2007) Spatial confusion or clarity? Reply to Ricotta & Avena. *J Veg Sci* 18:921–924
- R Development Core Team (2012) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. www.R-project.org/
- Ricotta C, Avena G (2006) On the evaluation of ordinal data with conventional multivariate procedures. *J Veg Sci* 17: 839–842
- Ricotta C, Feoli E (2013) Does ordinal cover estimation offer reliable quality data structures in vegetation ecological studies? *Folia Geobot* 48:437–447
- Robblee MB, Barber TR, Carlson PR Jr, Durako MJ and others (1991) Mass mortality of the tropical seagrass *Thalassia testudinum* in Florida Bay (USA). *Mar Ecol Prog Ser* 71:297–299
- Rose CD, Sharp WC, Kenworthy WJ, Hunt JH and others (1999) Overgrazing of a large seagrass bed by the sea urchin *Lytechinus variegatus* in Outer Florida Bay. *Mar Ecol Prog Ser* 190:211–222
- Tüxen R, Ellenberg H (1937) *Der systematische und ökologische Gruppenwert. Ein Beitrag zur Begriffsbildung und Methodik der Pflanzensoziologie*. Mitt Flor-Soz Arbeitsgem Niedersachsen 3:171–184
- Underwood AJ (1997) *Experiments in ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press, Cambridge
- van der Maarel E (1979) Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* 39:97–114
- van der Maarel E (2005) *Vegetation ecology*. Blackwell, Malden, MA
- van der Maarel E (2007) Transformation of cover-abundance values for appropriate numerical treatment: alternatives to the proposals by Podani. *J Veg Sci* 18:767–770
- Watling L, Kinner PC, Maurer D (1978) Use of species abundance estimates in marine benthic studies. *J Exp Mar Biol Ecol* 35:109–118

Editorial responsibility: Antony Underwood, Sydney, New South Wales, Australia

*Submitted: November 7, 2017; Accepted: April 16, 2018
Proofs received from author(s): June 2, 2018*