

The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling

Eric Dexter ,^{*1} Gretchen Rollwagen-Bollens ,^{1,2} Stephen M. Bollens^{1,2}

¹School of the Environment, Washington State University, Vancouver, Washington

²School of Biological Sciences, Vancouver, Washington

Abstract

Nonmetric multidimensional scaling (NMDS) is a powerful statistical tool which enables complex multivariate data sets to be visualized in a reduced number of dimensions. Users typically evaluate the fit of an NMDS ordination via ordination “stress” (i.e., data distortion) against a commonly accepted set of heuristic guidelines. However, these guidelines do not account for the mathematical relationship which links ordination stress to sample size. Consequently, researchers working with large data sets may unnecessarily present ordinations in an intractable number of dimensions, subdivide their data, or forego the use of NMDS entirely and lose the benefits of this highly flexible and useful technique. In order to overcome the limitations of these practices, we advocate for an alternative approach to the evaluation of NMDS ordination fit via the usage of permutation-based ecological null models. We present the rationale for this approach from a theoretical basis, supported by a brief literature review, and an example usage of the methodology. Our literature review shows that NMDS analyses often far exceed the number of observations under which the original stress guidelines were formulated—with a significant increasing trend in recent decades. Adoption of a permutation-based approach will consequently provide a more flexible and quantitative evaluation of NMDS fit and allow for the continued application of NMDS in an era of increasingly large datasets.

The trouble with stress

Researchers across a diverse array of fields are often tasked with teasing out hidden patterns of structure from within expansive and highly-dimensional datasets. Like many of our peers in ecology and oceanography, our working group often employs nonmetric multidimensional scaling (NMDS) as the method of choice for visualizing patterns of community abundance in a tractable number of dimensions—typically 2 or 3. NMDS is an extremely flexible technique for analyzing many different types of data, especially highly-dimensional data that exhibit strong deviations from assumptions of normality. Unlike some other ordination techniques, the NMDS approach fits data to a number of axes that are determined a priori to the analysis and does not contain hidden axes of variation. NMDS analysis can only be achieved through a computationally-dense (and somewhat opaque) algorithm that cannot be performed without the aid of a computer. Consequently, usage is heavily guided by heuristic guidelines and common practices within the field. These common practices guide users to the

minimum number of dimensions in which a given dataset can be visualized without inducing unacceptable levels of distortion (measured as ordination “stress”).

In preparing some data for a recent publication, we encountered a situation that led us to question these common practices. Specifically, this occurred when we attempted to visualize some plankton community data in a 2-dimensional NMDS ordination, and the resultant value of ordination stress (the *de facto* metric of ordination fit) exceeded the commonly accepted limit of 0.2 (or 20—depending on how the chosen software scales the stress values) (Kruskal 1964a; Clarke 1993; McCune and Grace 2002). As reviewers of earlier publications have insisted to us that this value be utilized as a strict cutoff, we reran the ordination seeking a 3-dimensional solution with lower stress. Although the resultant ordination achieved a stress value of less than 0.2, the addition of a third dimension provided little to no improvement in interpretative power, and in many ways obscured the major trends in the data.

As we consulted the ecological literature, we observed that rigid adherence to this threshold is not uncommon. We found that this threshold was commonly delineated in the methodological details of manuscripts (e.g., see Boyra et al. 2004; Vizzini and Mazzola 2004; Bowen et al. 2005; Heino

*Correspondence: eric.dexter@wsu.edu

Additional Supporting Information may be found in the online version of this article.

and Mykrä 2006; Righi-Cavallaro et al. 2010; White et al. 2011), and that users sometimes increased ordination dimensionality for the explicit purpose of achieving stress values below this threshold of 0.2 (Eallonardo and Leopold 2014; Breckenridge et al. 2015). Characteristic statements from the literature include "...only NMDS ordinations with a stress of 0.2 or lower were retained. Stress values >0.2 are generally considered poor and potentially uninterpretable" (Tyler and Kowalewski 2014) and "If the stress levels were greater than 0.2, the plots were considered difficult to interpret" (Boyra et al. 2004). Accordingly, we asked ourselves from where did these guidelines for stress arise, and what benefit is gained by adhering so rigidly to this cutoff value of 0.2?

In the following pages we aim to show that the generally accepted guidelines of ordination stress arose from the practical experience of early workers, and that these guidelines have become codified among some users in a manner likely unintended by the progenitors of the guidelines. Secondly, we demonstrate that sample sizes reported in the ecological and oceanographic literature now frequently exceed the conditions under which the original stress guidelines were developed, with a significant increasing trend in recent decades. Next, we outline an argument against the usage of any single set of fixed guidelines for ordination stress, supported by examples drawn from real and simulated ecological data. Finally, we highlight an easily-implemented alternative approach to the evaluation of NMDS ordination fit, based upon the employment of permutation-based null models of community structure.

A brief history of nonmetric multidimensional scaling and ordination stress

NMDS originated in the field of mathematical psychology as a method of ordination with inherently nonmetric variables such as personality traits and degrees of agreement/disagreement (Shepard 1980). Although development of the NMDS algorithm was somewhat of an iterative process (Shepard 1980), the core NMDS algorithm was presented in a pair of companion papers published in a 1962 issue of *Psychometrika* (Shepard 1962a,b). Two years later, J.B. Kruskal published a second set of papers in *Psychometrika* which expanded upon and substantially improved the performance of the NMDS algorithm (Kruskal 1964a,b).

Because NMDS produces an easily interpreted output and is relatively free of assumptions about distributions of data, the method has found broad application across a number of fields (Kruskal and Wish 1978; Borg and Groenen 2005). Initial ecological applications of NMDS were largely restricted to the study of terrestrial plant communities and tended to be formidably complex in their technical details (Anderson 1971, Prentice 1977, Matthews 1978). In the ensuing years, a number of authors directly compared the performance of various ordination methods in recovering patterns of ecological

structure from simulated community datasets, and generally found NMDS to be the most robust approach in most situations (Fasham 1977; Kenkel and Orloci 1986; Minchin 1987). However, ecological applications of NMDS largely remained restricted to terrestrial plant communities until the publication of several key papers in the 1980's which outlined a general strategy for nonmetric multivariate analysis of community data (e.g., Field et al. 1982; Kenkel and Orloci 1986). Subsequent advances in computing power and the widespread availability of statistical software tools for NMDS have dramatically lowered the technical barriers to use of this algorithm, and as such, this formerly esoteric methodology is now often taught in graduate-level surveys of multivariate statistics (McCune and Grace 2002; Zuur et al. 2007).

Returning to the initial development of the NMDS algorithm, the key innovation of Kruskal's implementation of NMDS relative to previous incarnations of the algorithm was the introduction of a quantitative measure of ordination fit—a measure which he defined as "stress" (Kruskal 1964a,b; Shepard 1980). Although stress is often discussed in terms of ordination fit (Kruskal and Carroll 1969), stress is actually the product of a normalized loss function which minimizes the dissimilarity between rank order distances (De Leeuw and Stoop 1984). This metric indicates how well the algorithm has managed to arrange the points in the ordination while preserving the rank-order distances (i.e., smallest distance, 2nd smallest distance, 3rd smallest distance, etc...) as represented in the original matrix—measured as deviation from a monotonically increasing (i.e., each value is greater than the previous) function (Kruskal 1964a). Although two slightly different calculations for stress exist, by far the most commonly utilized measure of stress is known as "stress 1" and is given as:

$$\text{Stress}_1^2 = \frac{\sum_{i,j} (d_{ij} - \tilde{d}_{ij})^2}{\sum_{i,j} d_{ij}^2}. \quad (\text{Equation 1})$$

Here the term d_{ij} represents the actual distance between samples i and j in ordination space, while \tilde{d}_{ij} represents the fitted distance (in respect to the monotonic regression line) between samples i and j . In the numerator, these differences are squared so that only positive values are returned and then summed across all pairwise combinations of samples. The denominator is simply the sum of each squared d_{ij} , which serves as a scaling factor. The lowest possible value of stress is 0, which indicates complete accordance between all rank order distances in the input data and the final ordination. Increasing amounts of discordance between the input and the ordination will be reflected in values of stress greater than 0. Although stress (Kruskal's first formulation) is nominally measured on a scale from 0 to 1, the maximum value of stress which can be attained in 2 dimensional solutions is approximately 0.58, and this upper limit

will decrease with increasing NMDS dimensionality (De Leeuw and Stoop 1984).

For the researcher, one difficulty lies in deciding when a value of stress is sufficiently high to caution against biological interpretation of an ordination. To address this issue, several authors (starting with J.B. Kruskal in his original 1964 manuscripts) have proposed guidelines for the interpretation of NMDS stress based upon a series of cutoff values. K.R. Clarke's highly-influential "Non-parametric multivariate analyses of changes in community structure" (1993) is the most commonly employed resource for these guidelines, with more than 8000 citations to date. Expanding upon recommendations in Kruskal's original NMDS papers, Clarke asserts (here slightly condensed):

"Stress <0.05 gives an excellent representation with no prospect of misinterpretation.

Stress <0.1 corresponds to a good ordination with no real risk of drawing false inferences...

Stress <0.2 can still lead to a usable picture, although for values at the upper end of this range there is potential to mislead...

Stress >0.2 is likely to yield plots which could be dangerous to interpret. Certainly by the time stress reaches 0.35-0.4 the samples are effectively randomly placed..."

It should be noted that although stress is often discussed in terms of goodness-of-fit (or badness-of-fit) it does not represent either in the conventional sense—but rather is an algorithm optimization criterion. In the same sense that a low Akaike Information Criteria (AIC) score does not imply that a given regression model is an accurate representation of a given system, a low stress value does not provide researchers with *carte blanche* to freely interpret an NMDS ordination. Clarke (1993) emphasized this point by following his set of guidelines with an explicit warning against over-reliance on stress, and pointing toward several complementary methods for evaluating ordination fit (e.g., Shepard plots, scree plots, cluster analysis, etc.). Irrespective of these warnings, the evaluation of stress against a ceiling of 0.2 (or a rescaled value of 20) appears to have become an important (or in some cases—primary) method of evaluating NMDS ordination fit for many users.

Our brief survey of the NMDS literature (see below) found that almost every article reported ordination stress values, while virtually none of the manuscripts mentioned the use of any other criteria to evaluate ordination fit. While a small number of these studies did present ordinations with stress values greater than 0.20, such examples were infrequent (Clarke and Warwick 1994; Olsgard and Gray 1995; Paredes et al. 2014; Gutow et al. 2015), and in some cases authors explicitly stated that they had taken steps to avoid crossing this threshold. For example Eallonardo and Leopold (2014) stated that "Community composition data were [shown by] plot scores from two-axis NMDS analyses (except for the *P.*

australis community which required three axes to achieve an NMDS stress level less than 20)".

Not all stress is created equal

As mentioned above, many users have tended to treat these guidelines as a series of concrete rules. However, there are several strong arguments against the use of these guidelines in such a manner, and moreover, against strict adherence to any single set of guidelines. Firstly, when the number of observations is small, the ordination tends to become over-fitted, which links increasing sample size with an increase in stress (Borg and Groenen 2005). Secondly, stress values tend to decrease with increasing ordination dimensionality, which complicates the comparison of stress values between ordinations with differing numbers of dimensions. The sample size effect on stress has been previously reported in Kruskal and Wish (1978) and graphically shown by McCune and Grace (2002, fig. 16.4), but given the continued overreliance of these stress cutoff values by some members of the research community it may be of some utility to provide a clear demonstration of these two properties.

An additional consideration before moving ahead to demonstration of these properties of stress is that the strategy utilized to deal with tied values in the dissimilarity matrix can strongly affect the final value of stress achieved in an NMDS ordination (Kruskal 1964a; McCune and Mefford 2016). In what Kruskal (1964a) termed the "primary" approach, ties in the dissimilarity matrix are permitted to correspond to distances in the ordination space that are not tied. Conversely, Kruskal's "secondary" approach constrains the NMDS algorithm such that ties in the dissimilarity matrix must correspond to tied distances in the ordination space. Consequently, the constraints imposed upon the NMDS algorithm by the secondary approach to handling ties may yield relatively larger values of stress than analyses using the primary approach. This point is particularly relevant to types of data that are likely to contain ties, such as presence-absence data, or matrices containing a small number of variables (McCune and Mefford 2016; Supporting Information Fig. S1 of this manuscript)

As users of NMDS typically don't report which approach was used to break ties, this ambiguity further compounds the difficulty of evaluating stress in absolute terms. Readers should be cautioned against making assumptions about the treatment of ties in any given work, as popular statistical software suites operate using different approaches. For example the MonoMDS engine implemented in the R package *vegan* uses the primary approach under default settings (Oksanen et al. 2017) while PC-ORD uses the secondary approach under default settings in most, but not all, releases (McCune and Mefford 2016).

A demonstration of the trouble with stress using field-derived and simulated data

At this point we will now introduce two types of data that will be used to illustrate several of the aforementioned points: field-derived and simulated ecological data. The field data are a subset from an ongoing time-series of zooplankton samples collected from the lower Columbia River in the Northwest United States. This dataset contains 103 consecutive monthly samples (representing ~ 8.5 yr) comprised of 35 zooplankton taxa, collected from a single fixed location near Vancouver, WA. This particular dataset will be utilized because the large number of observations can be readily subsampled to elucidate the effect of sample size on ordination output, and because an extensive NMDS-based analysis of these data is available for readers who seek further biological context (Dexter et al. 2015).

In order to generalize our conclusions beyond this single dataset, and indeed, to show that the patterns which we wish to illustrate can be demonstrated across datasets of different dimensions, a large number of simulated ecological datasets were generated and analyzed in parallel with the Columbia River zooplankton data. This approach is consistent with the historical NMDS literature, which has often employed ordination of simulated ecological datasets to demonstrate specific properties of the algorithm (Fasham 1977; Kenkel and Orloci 1986; Minchin 1987).

The simulated community data were structured across two equally weighted ecological gradients and comprised of 10–100 hypothetical species. The total abundance of each species was normally distributed across the two ecological gradients, with each species' maximum abundance centered on a randomly assigned niche optima. The width of each niche optimum was also randomly assigned such that generalist species would be distributed across the entire simulated landscape, specialists would be concentrated within a localized region, and other species would exhibit distributional patterns intermediate between these two extremes. From this virtual landscape 5–100 samples were randomly drawn. In order to better approximate stochastic sampling-effects, individual species counts were determined by sampling from a negative binomial distribution with the shape parameters of the distribution based upon the niche optimum value for that particular species at that particular location within the simulated landscape. We acknowledge that these simulations represent only rough approximations of ecological processes, but offer that the resultant data exhibit a reasonable level of heterogeneity and complexity for the purposes at hand. All ecological simulations were conducted using the *coenocliner* package v0.2-2 (Simpson et al. 2016) in R v3.2.2 (R Core Team 2015).

Similarly, the Columbia River zooplankton data were repeatedly subsampled to sets of 5–100 observations, ordinated via NMDS, and summarized by the final stress value achieved. This entire process was repeated across 2, 3, and 4 dimensions of ordination space, with the resulting mean

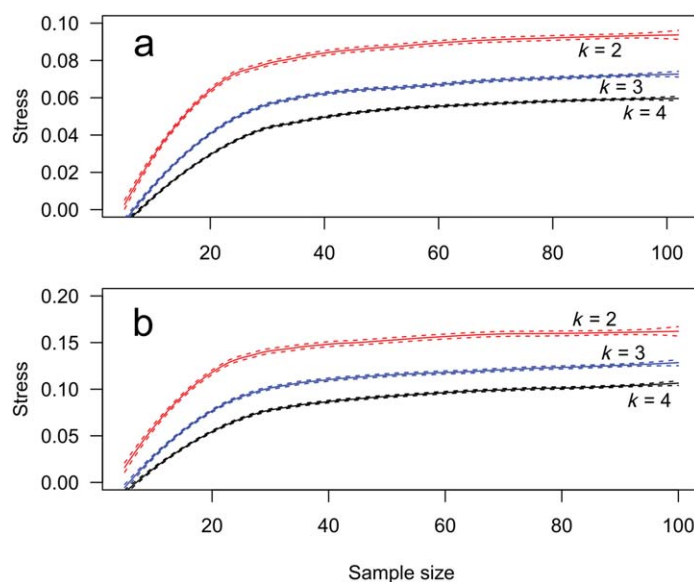


Fig. 1. The asymptotically increasing relationship between ordination stress and sample size as shown by (a) ordination of subsamples from the Columbia River zooplankton series and (b) ordination of independently generated simulated ecological datasets. Each panel plots this relationship for ordination in ($k=2,3,4$) dimensions, showing a clear decrease in ordination stress with increasing values of k . Each of the six data series are comprised of 10 independently generated samples at each value of n , with the resulting mean (\pm SE) shown as a loess smoothed line with span set to 0.5.

(\pm SE) stress at each n shown as a loess smoothed regression line with span = 0.5 (Fig. 1a). 10 independent subsets were ordinated at each value of n for each series, resulting in a total of 950 ordinations per series. The analysis of the simulated ecological data varied only in the respect that simulated datasets were generated independently, rather than through the subsetting of a larger dataset. In all other respects the analyses and associated conclusions were unchanged between field-derived and simulated ecological data (Fig. 1b).

The results of these experiments broadly agree with patterns shown by Kruskal and Wish (1978) and McCune and Grace (2002, fig. 16.4), and clearly demonstrate that stress increases with increasing sample size and decreases with increasing ordination dimensionality, essentially irrespective of the underlying data. We can thus conclude that the commonly employed stress guidelines should not be applied to NMDS ordinations of greater than 2 dimensions, and that even within 2-dimensional solutions, the guidelines become less reliable as studies diverge from the range of sample sizes under which they were developed. Furthermore, we can see that this sample size effect is particularly acute when the number of samples is less than 20, but becomes greatly diminished as the number of samples approaches 30–40.

This latter point is especially critical given that datasets appear to be increasing in size over time—as one might

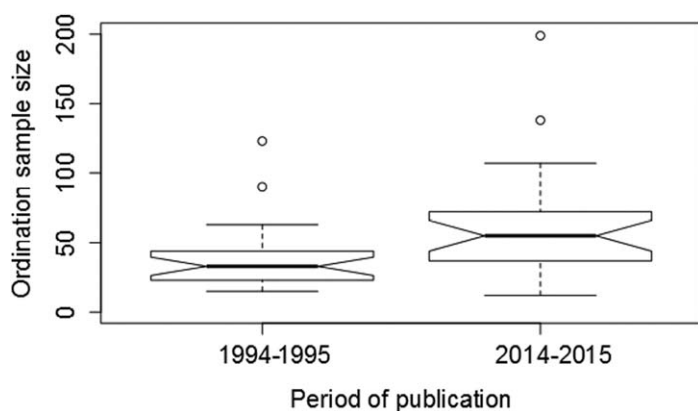


Fig. 2. The increasing trend in NMDS ordination samples sizes as observed in our review of the ecological and oceanographic literature. There is a statistically significant ($p = 0.0298$; two-tailed unpaired t -test with Welches correction) increase in sample size across the time periods, with mean sample size of 38.8 ± 5.0 during 1994–1995 ($n = 25$) and 60.5 ± 8.3 during 2014–2015 ($n = 25$).

expect given the ongoing maturation of long-term studies and the growing prevalence of multi-investigator projects. Indeed, our research group has published many papers ($n = 13$) using NMDS, but only in relatively recent years have our sample sizes grown to $n > 100$ (Bollens et al. 2014; Dexter et al. 2015; Lee et al. 2015; Hassett et al. 2017).

To highlight this issue, we compared 25 randomly selected articles in the ecological and oceanographic literature that utilized NMDS and were published in 1994–1995 (just after Clarke’s highly influential manuscript was published) against the same number of randomly selected articles published in 2014–2015. We found a statistically significant ($p = 0.0298$; two-tailed unpaired t -test with Welches correction) increase in sample size across this period, with mean sample size of 38.8 ± 5.0 during 1994–1995 and 60.5 ± 8.3 during 2014–2015 (Fig. 2).

Similar issues have arisen in the application of NMDS to genomic data, as modern genomic studies may contain thousands of individual markers examined across highly dimensional ($k > 10$) ordinations (Tzeng et al. 2008; Jakaitiene et al. 2016). Approaches to surmounting issues of sample size in this field have been highly variable, and sometimes involve a complete reworking of the core NMDS algorithm (Taguchi and Oono 2005; Tzeng et al. 2008), or multiple-stage approaches which incorporate additional ordination methods (Zhu and Yu 2009). Clearly, the wider research community would also benefit from a more flexible framework to evaluate NMDS stress across a larger range of sample sizes and higher dimensional spaces.

Toward a better interpretation of stress

Efforts to devise a more flexible approach to the evaluation of ordination stress began to appear only a few years

following Kruskal’s 1964 formulation of NMDS (Klahr 1969; Stenson and Knoll 1969), and in the ensuing decades, researchers harnessed increasingly powerful computers to update and extend this framework (Spence and Ogilvie 1973; Levine 1978; MacCallum 1981; De Leeuw and Stoop 1984). This flexible approach is based upon the repeated ordination of random matrices in order to estimate the distribution of stress values that arise from unstructured data, given a specific sample size and dimensionality. Such an approach erects a specific null hypothesis (data are randomly structured) which is either rejected or accepted via the consultation of probability tables published for this specific purpose.

Due to the massive computational power required to produce thousands of NMDS ordinations, a truly robust set of tables were not produced until the beginning of the 21st century (Sturrock and Rocha 2000). The tables generated by Sturrock and Rocha (which remain the most exhaustive to date) are derived from over 500,000 randomly generated matrices, each composed of 4–100 samples of 1–10 variables. These easily-interpreted probability tables provide mean stress values and 95% confidence intervals for datasets containing up to 100 observations in 1, 2, and 3 dimensions (Sturrock and Rocha 2000). Accordingly, the evaluation of NMDS stress against these probability tables has begun to gain traction across a wide range of fields, including anthropology (Handwerker 2002), population genetics (Urbach et al. 2007), human genomics (Batini et al. 2011), and linguistics (Wnuk and Majid 2014). There remains, however, a major flaw in the formulation of this approach that hinders application of this approach for many types of data—namely, the formulation of the null model. Sturrock and Rocha (2000), for example, populated their matrices by randomly drawing values from a uniform distribution. If one considers this random matrix as an ecological dataset, a series of problems become readily apparent. In these data, all species are present in all samples, all species are equally abundant, and all samples are equally diverse. In other words, this particular null model (i.e., random structure) would tend toward rejection when evaluated against virtually any ecological data. This conceptual flaw is not fatal, however, and is easily remedied through the employment of a more ecologically appropriate null model (with an important caveat described below).

The generation of random matrices via a Poisson, negative binomial, or zero-inflated distribution would produce a much better approximation of ecological data from which probability tables could be derived. Given the computing power readily available to modern researchers, it would be a relatively simple task to recreate the probability tables developed by Sturrock and Rocha (2000) (which were not specifically intended for ecological applications) under a more ecologically plausible framework. However, such an approach would still prove unsatisfactory in a number of ways. Most crucially, stress is also dependent upon the

number of variables in any one sample, and that for groups which share no overlapping members the stress is essentially undefined (Shepard 1980; Borg and Groenen 2005). As such, ecologically defensible stress probability tables would need to allow for variation in the size of the total species pool and the rarity of individual species—essentially turning a 2-dimensional lookup table into an impractically complex set of tables.

We believe a more promising approach for the evaluation of stress lays not in the consultation of pre-made lookup tables, but rather via iterative permutation of the dataset of interest. We are not alone in this enthusiasm for permutation-based approaches, as the basic idea of using permutation tests to evaluate NMDS ordinations has been around in some form or another for at least two decades. In particular, our proposal builds upon and extends a Monte Carlo randomization approach advocated by McCune and Grace (2002) and implemented in the PC-ORD statistical software (McCune and Mefford 2016). We aim here to elucidate and extend the basic conceptual framework underlying such permutation-based approaches, and to demonstrate the manner by which such approaches may be employed.

Briefly summarized, we advocate for the evaluation of NMDS stress under an ecological null model framework. At the most fundamental level, ecological null-models are a permutation-based approach to evaluating specific biological hypotheses against a background of heterogeneous data. As defined in *Null Models in Ecology* (Gotelli and Graves 1996), a null model is “a pattern-generating model that is based on the randomization of ecological data or random sampling from a known or imagined distribution... Certain elements of the data are held constant, and others are allowed to vary stochastically to create new assemblage patterns. The randomization is designed to produce a pattern that would be expected in the absence of a particular ecological mechanism.” In this instance, the specific null model to be tested is that permutations of the underlying dataset can produce equivalent stress values to those achieved by the original data. In ecological terms, failure to reject the null hypothesis would indicate that a given ordination structure reflected stochastic sampling of heterogeneously distributed species rather than a systematic pattern of species-associations.

Of course, a potential pitfall of this approach is that the permutation algorithm may be misspecified so that it does not incorporate realistic biological constraints and produces a null model which is too easily rejected (Gotelli and Graves 1996). This was the aforementioned pitfall encountered when sampling from a random uniform distribution. Indeed the possibility for incorrect specification of the null model lies at the center of vociferous criticism of the early implementations of these models (Grant and Abbott 1980; Diamond and Gilpin 1982; Quinn and Dunham 1983; Roughgarden 1983), and spurred considerable theoretical and methodological refinements to the use of ecological null

models (Gotelli 2001; Gotelli and Entsminger 2003; Gotelli and Ulrich 2012). Fortunately, researchers are not forced to construct and evaluate permutation algorithms for null model construction from scratch, as there exists a large and robust body of literature on this topic, which is well integrated into the available software tools for ecological null model construction (e.g., Picante—Kembel et al. 2014; EcoSimR—Gotelli et al. 2015; Vegan—Oksanen et al. 2017).

Example application of the ecological null model approach to evaluating NMDS stress

As an example, we will demonstrate the straightforward manner in which this permutation-based approach can be employed to evaluate NMDS stress using three sets of data: the aforementioned Columbia River zooplankton series (Dexter et al. 2015), a simulated biological community exhibiting a moderate degree of community structure (i.e., strong associations between some species), and a simulated community exhibiting a low degree of community structure. The varying degrees of community structure among the simulated datasets were obtained by altering the niche breadth of individual species, such that the weakly structured community is exclusively comprised of generalist species.

Each of the three datasets were evaluated against a null hypothesis of unstructured species associations through NMDS ordination of 1000 independent permutations of each dataset. For each dataset, we tested the hypothesis that community composition (and thus placement of points on the ordination) is lacking in systematic structure. In each case, the null hypothesis was rejected if p -values less than 0.05 were obtained when employing a one sample z -test. Note that the z -distribution is a reasonable approximate of the t -distribution because we are evaluating $n = 1000$ permutations, but this approximation would not hold for low values of n . The permutation algorithm was set to retain the total counts per sampling unit, total counts per species, and the number of zero-cells in the whole dataset. Although this procedure may appear cumbersome, all steps of our analysis—ordination, permutation (algorithm = “quasiSwap count”), and test statistic calculation—were achieved through a single call of the *Oecosimu* function found in the *vegan* package v2.4-3 for R (R Core Team 2015; Oksanen et al. 2017). All NMDS analyses presented in this manuscript were conducted using Bray–Curtis dissimilarities derived from untransformed species abundances, and with ties in the dissimilarity matrix treated according to Kruskal’s primary approach (no penalties for ties).

As to the results of this analysis, Fig. 3a (Columbia River zooplankton) shows clear evidence in favor of rejecting the null hypothesis of unstructured community data ($z = -5.51$; $p < 0.001$). When combined with complementary metrics (e.g., clustering algorithms, Shepard’s plot, etc.), this result provides a strong basis for biological interpretation of the

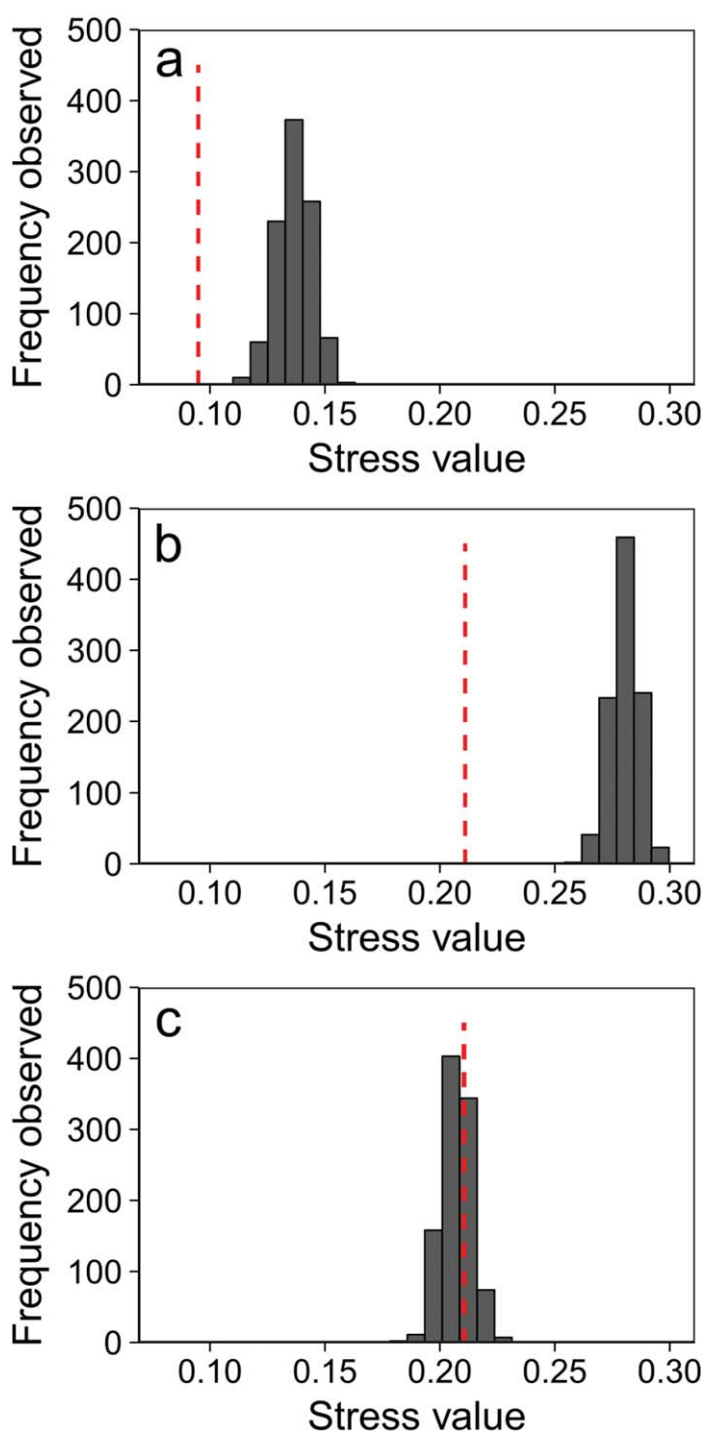


Fig. 3. The distribution of stress values achieved via permutation of (a) Columbia River zooplankton data, (b) simulated ecological data with moderate community structure, and (c) simulated ecological data with weak community structure. The stress value achieved for ordination of each raw dataset is shown as the vertical dashed line in each plot. There is sufficient evidence to reject the ecological null hypothesis (no systematic community structure) for the Columbia River zooplankton data ($z = -5.51$; $p < 0.001$) and the moderately structured simulation data ($z = -11.27$; $p < 0.001$), but not the weakly structured simulation data ($z = 0.47$; $p = 0.671$).

ordination. Figure 3b (moderately structured simulation data) also shows clear evidence in favor of rejecting the null hypothesis ($z = -11.27$; $p < 0.001$). Note that the stress value associated with this set of simulated data exceeds the conventional stress cutoff of 0.20, yet the results of the permutation approach indicate that the ordination contains a meaningful degree of structure (which indeed was the programmed behavior of the simulation). In contrast, Fig. 3c (weakly structured simulation data) shows insufficient evidence to reject the null hypothesis ($z = 0.47$; $p = 0.671$), which cautions strongly against any attempt to infer biological conclusions from the ordination of these data. Of particular note is the fact that the stress values achieved in both sets of simulations were roughly equivalent (and greater than 0.2), but the permutation-based approach clearly differentiated the systematically structured communities from those structured only by stochastic sampling artifacts.

In summary, we have evaluated the NMDS ordination fit for each of the three different data series, using cutoff values appropriately matched to the sample size, ordination dimensionality, and underlying level of heterogeneity of each dataset. This approach can be distilled into a small number of steps:

1. Define a specific ecological null hypothesis to be tested. For the purpose of evaluating NMDS stress in biological community data, a reasonable hypothesis would be that species associations are random.
2. Choose a permutation algorithm appropriately suited to the null hypothesis and dataset of interest.
3. Conduct NMDS ordination on many independent permutations of the original dataset and record the values of stress associated with each permutation. We have found that 1000 iterations tend to be more than sufficient.
4. Evaluate the stress value associated with the original dataset against the distribution of stress values from the permuted datasets using a one-sample z-test. By convention, we have chosen to set alpha at 0.05 using a two-tailed test.
5. If there is sufficient evidence to reject the null hypothesis, and associated metrics of ordination fit (Shepard's plots, clustering, etc.) are likewise satisfactory, proceed with efforts to produce a biological interpretation of NMDS ordination.

Conclusion

Kruskal (1964a), Clarke (1993), Sturrock and Rocha (2000), and many other authors have clearly stated that stress should not be used as the sole criterion for evaluation of NMDS ordination fit—with which we agree. Nonetheless, the appeal of a hard numerical cutoff appears to have favored the use of 0.2 as a firm limit among some users of NMDS. This unintended rule becomes entirely untenable

when we consider that stress values cannot be readily compared across different levels of dimensionality, data structure (e.g. binary vs. count), tie handling strategies, nor between datasets of greatly different size.

Probability-based approaches to the evaluation of ordination fit have gained considerable traction in some fields of research as a more flexible framework for the evaluation of stress, but in their present form lack an appropriate null model for the evaluation of biological community data. The ecological null model approach of constrained permutation demonstrated here provides an easily implemented method for generating an appropriate null model for evaluating NMDS ordination structure against a background of heterogeneous community data.

Although we have devoted considerable space to the discussion of NMDS stress, we wish to reiterate that stress is not a stand-alone measure of NMDS interpretability, but rather a metric which directs the NMDS optimization algorithm. The methodological approach we propose here allows one to evaluate an NMDS ordination against a specific (but useful) null hypothesis, but it does not provide any information on the appropriateness of the chosen number of dimensions, or the fit of individual points in the ordination space. For a fuller overview of such aspects of modern NMDS methodology we highly recommend McCune and Grace (2002), Zuur et al. (2007) or any trusted textbook on multivariate statistical analysis, while those interested in more technical aspects of the methodology are recommended to consult Borg and Groenen (2005) as well as the aforementioned foundational NMDS papers.

Investigators wishing to employ this permutation-based approach will find a wealth of freely available software tools which can be easily integrated into existing workflows. The vegan package for R (Oksanen et al. 2017) was used in the analysis presented here, but other software tools (for example PC-ORD) are available to perform similar functions. For those wishing a more detailed tutorial, all analyses presented in this workflow are available as a supplemental resource to this article in the form of clearly annotated R script files.

References

- Batini, C., and others. 2011. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Mol. Biol. Evol.* **28**: 1099–1110. doi:[10.1093/molbev/msq294](https://doi.org/10.1093/molbev/msq294)
- Bollens, S. M., J. K. Breckenridge, J. R. Cordell, C. A. Simenstad, and O. Kalata. 2014. Zooplankton of tidal marsh channels in relation to environmental variables in the upper San Francisco Estuary. *Aquat. Biol.* **21**: 205–219. doi:[10.3354/ab00589](https://doi.org/10.3354/ab00589)
- Borg, I., and P. Groenen. 2005. Modern multidimensional scaling - theory and applications. Springer.
- Bowen, K. D., R.-J. Spencer, and F. J. Janzen. 2005. A comparative study of environmental factors that affect nesting in Australian and North American freshwater turtles. *J. Zool.* **267**: 397–404. doi:[10.1017/S0952836905007533](https://doi.org/10.1017/S0952836905007533)
- Boyra, A., P. Sanchez-Jerez, F. Tuya, F. Espino, and R. Haroun. 2004. Attraction of wild coastal fishes to an Atlantic Subtropical Cage Fish Farms, Gran Canaria, Canary Islands. *Environ. Biol. Fishes* **70**: 393–401. doi:[10.1023/B:EBFI.0000035435.51530.c8](https://doi.org/10.1023/B:EBFI.0000035435.51530.c8)
- Breckenridge, J. K., S. M. Bollens, G. Rollwagen-Bollens, and G. C. Roegner. 2015. Plankton assemblage variability in a river-dominated temperate estuary during late spring (High-flow) and late summer (low-flow) periods. *Estuar. Coasts* **38**: 93–103. doi:[10.1007/s12237-014-9820-7](https://doi.org/10.1007/s12237-014-9820-7)
- Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**: 117–143. doi:[10.1111/j.1442-9993.1993.tb00438.x](https://doi.org/10.1111/j.1442-9993.1993.tb00438.x)
- Clarke, K., and R. Warwick. 1994. Similarity-Based testing for community pattern - the 2-way layout with. *Mar. Biol.* **118**: 167–176. doi:[10.1007/BF00699231](https://doi.org/10.1007/BF00699231)
- De Leeuw, J., and I. Stoop. 1984. Upper bounds for Kruskal's stress. *Psychometrika* **49**: 391–402. doi:[10.1007/BF02306028](https://doi.org/10.1007/BF02306028)
- Dexter, E., S. M. Bollens, G. Rollwagen-Bollens, J. Emerson, and J. Zimmerman. 2015. Persistent vs. ephemeral invasions: 8.5 years of zooplankton community dynamics in the Columbia River. *Limnol. Oceanogr.* **60**: 527–539. doi:[10.1002/lno.10034](https://doi.org/10.1002/lno.10034)
- Diamond, J. M., and M. E. Gilpin. 1982. Examination of the “null” model of connor and simberloff for species co-occurrences on Islands. *Oecologia* **52**: 64–74. doi:[10.1007/BF00349013](https://doi.org/10.1007/BF00349013)
- Eallonardo, A. S., and D. J. Leopold. 2014. Inland Salt Marshes of the Northeastern United States: Stress, disturbance and compositional stability. *Wetlands* **34**: 155–166. doi:[10.1007/s13157-013-0493-y](https://doi.org/10.1007/s13157-013-0493-y)
- Fasham, M. J. R. 1977. A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines, and coenoplanes. *Ecology* **58**: 551–561. doi:[10.2307/1939004](https://doi.org/10.2307/1939004)
- Field, J. G., K. R. Clarke, and R. M. Warwick. 1982. A practical strategy for analysing multispecies distribution patterns. *Mar. Ecol. Prog. Ser.* **8**: 37–52. doi:[10.3354/meps008037](https://doi.org/10.3354/meps008037)
- Gotelli, N. J. 2001. Research frontiers in null model analysis. *Glob. Ecol. Biogeogr.* **10**: 337–343. doi:[10.1046/j.1466-822X.2001.00249.x](https://doi.org/10.1046/j.1466-822X.2001.00249.x)
- Gotelli, N. J. and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press.
- Gotelli, N. J., and G. L. Entsminger. 2003. Swap algorithms in null model analysis. *Ecology* **84**: 532–535. doi:[10.1890/0012-9658\(2003\)084\[0532:SAINMA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0532:SAINMA]2.0.CO;2)
- Gotelli, N. J., and W. Ulrich. 2012. Statistical challenges in null model analysis. *Oikos* **121**: 171–180. doi:[10.1111/j.1600-0706.2011.20301.x](https://doi.org/10.1111/j.1600-0706.2011.20301.x)

- Gotelli, N., E. Hart, and A. Ellison. 2015. EcoSimR: Null Model Analysis for Ecological Data.
- Grant, P. R., and I. Abbott. 1980. Interspecific competition, island biogeography and null hypotheses. *Evolution* **34**: 332–341. doi:[10.2307/2407397](https://doi.org/10.2307/2407397)
- Gutow, L., J. Beermann, C. Buschbaum, M. M. Rivadeneira, and M. Thiel. 2015. Castaways can't be choosers — Homogenization of rafting assemblages on floating seaweeds. *J. Sea Res.* **95**: 161–171. doi:[10.1016/j.seares.2014.07.005](https://doi.org/10.1016/j.seares.2014.07.005)
- Handwerker, W. P. 2002. The construct validity of cultures: Cultural diversity, culture theory, and a method for ethnography. *Am. Anthropol.* **104**: 106–122. doi:[10.1525/aa.2002.104.1.106](https://doi.org/10.1525/aa.2002.104.1.106)
- Hassett, W., S. M. Bollens, T. D. Counihan, G. Rollwagen-Bollens, J. Zimmerman, S. Katz, and J. Emerson. 2017. Veligers of the invasive Asian clam *Corbicula fluminea* in the Columbia River Basin: broadscale distribution, abundance, and ecological associations. *Lake Reserv. Manag.* **33**: 234–248. doi:[10.1080/10402381.2017.1294218](https://doi.org/10.1080/10402381.2017.1294218)
- Heino, J., and H. Mykrä. 2006. Assessing physical surrogates for biodiversity: Do tributary and stream type classifications reflect macroinvertebrate assemblage diversity in running waters?. *Biol. Conserv.* **129**: 418–426. doi:[10.1016/j.biocon.2005.11.009](https://doi.org/10.1016/j.biocon.2005.11.009)
- Jakaitiene, A., M. Sangiovanni, M. R. Guarracino, and P. M. Pardalos. 2016. Multidimensional scaling for genomic data, p. 129–139. *In* Advances in stochastic and deterministic global optimization. Springer.
- Kembel, S. W., and others. 2014. picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464. doi:[10.1093/bioinformatics/btq166](https://doi.org/10.1093/bioinformatics/btq166)
- Kenkel, N., and L. Orloci. 1986. Applying metric and non-metric multidimensional-scaling to ecological-studies - some new results. *Ecology* **67**: 919–928. doi:[10.2307/1939814](https://doi.org/10.2307/1939814)
- Klahr, D. 1969. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. *Psychometrika* **34**: 319–330. doi:[10.1007/BF02289360](https://doi.org/10.1007/BF02289360)
- Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**: 1–27. doi:[10.1007/BF02289565](https://doi.org/10.1007/BF02289565)
- Kruskal, J. B. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**: 115–129. doi:[10.1007/BF02289694](https://doi.org/10.1007/BF02289694)
- Kruskal, J. B., and J. D. Carroll. 1969. Geometric models and badness-of-fit functions, *In* Multivariate analysis II. Academic Press.
- Kruskal, J. B. and M. Wish. 1978. Multidimensional scaling. SAGE.
- Lee, T. A., G. Rollwagen-Bollens, and S. M. Bollens. 2015. The influence of water quality variables on cyanobacterial blooms and phytoplankton community composition in a shallow temperate lake. *Environ. Monit. Assess.* **187**: 315. doi:[10.1007/s10661-015-4550-2](https://doi.org/10.1007/s10661-015-4550-2)
- Levine, D. M. 1978. A Monte Carlo study of Kruskal's variance based measure on stress. *Psychometrika* **43**: 307–315. doi:[10.1007/BF02293641](https://doi.org/10.1007/BF02293641)
- MacCallum, R. 1981. Evaluating goodness of fit in nonmetric multidimensional scaling by ALSCAL. *Appl. Psychol. Meas.* **5**: 377–382. doi:[10.1177/014662168100500311](https://doi.org/10.1177/014662168100500311)
- McCune, B. and J. B. Grace. 2002. Analysis of ecological communities. MjM Software Design.
- McCune, B., and M. J. Mefford. 2016. PC-ORD. Multivariate analysis of ecological data. MjM Software Design.
- Minchin, P. R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* **69**: 89–107. doi:[10.1007/BF00038690](https://doi.org/10.1007/BF00038690)
- Oksanen, J., and others. 2017. vegan: Community Ecology Package.
- Olsgard, F., and J. S. Gray. 1995. A comprehensive analysis of the effects of offshore oil and gas exploration and production on the benthic communities of the Norwegian continental shelf. *Mar. Ecol. Prog. Ser.* **122**: 277–306. doi:[10.3354/meps122277](https://doi.org/10.3354/meps122277)
- Paredes, M. A., V. Montecino, V. Anic, M. Egaña, and L. Guzmán. 2014. Diatoms and dinoflagellates macroscopic regularities shaped by intrinsic physical forcing variability in Patagonian and Fuegian fjords and channels (48°–56°S). *Prog. Oceanogr.* **129**: 85–97. doi:[10.1016/j.poccean.2014.07.002](https://doi.org/10.1016/j.poccean.2014.07.002)
- Quinn, J. F., and A. E. Dunham. 1983. On hypothesis testing in ecology and evolution. *Am. Nat.* **122**: 602–617. doi:[10.1086/284161](https://doi.org/10.1086/284161)
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Righi-Cavallaro, K. O., K. F. Roche, O. Froehlich, and M. R. Cavallaro. 2010. Structure of macroinvertebrate communities in riffles of a Neotropical karst stream in the wet and dry seasons. *Acta Limnol. Bras.* **22**: 306–316. doi:[10.4322/actalb.02203007](https://doi.org/10.4322/actalb.02203007)
- Roughgarden, J. 1983. Competition and theory in community ecology. *Am. Nat.* **122**: 583–601. doi:[10.1086/284160](https://doi.org/10.1086/284160)
- Shepard, R. 1980. Multidimensional-scaling, tree-fitting, and clustering. *Science* **210**: 390–398. doi:[10.1126/science.210.4468.390](https://doi.org/10.1126/science.210.4468.390)
- Shepard, R. N. 1962a. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**: 125–140. doi:[10.1007/BF02289630](https://doi.org/10.1007/BF02289630)
- Shepard, R. N. 1962b. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* **27**: 219–246. doi:[10.1007/BF02289621](https://doi.org/10.1007/BF02289621)
- Simpson, G. L., J. Oksanen, and F. Rodriguez-Sanchez. 2016. coenocliner: Coenocline simulation.
- Spence, I., and J. C. Ogilvie. 1973. A table of expected stress values for random rankings in nonmetric multidimensional scaling. *Multivar. Behav. Res.* **8**: 511–517. doi:[10.1207/s15327906mbr0804_8](https://doi.org/10.1207/s15327906mbr0804_8)

- Stenson, H., and R. Knoll. 1969. Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychol. Bull.* **71**: 122. doi:[10.1037/h0026864](https://doi.org/10.1037/h0026864)
- Sturrock, K., and J. Rocha. 2000. A multidimensional scaling stress evaluation table. *Field Methods* **12**: 49–60. doi:[10.1177/1525822X0001200104](https://doi.org/10.1177/1525822X0001200104)
- Taguchi, Y., and Y. Oono. 2005. Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics* **21**: 730–740. doi:[10.1093/bioinformatics/bti067](https://doi.org/10.1093/bioinformatics/bti067)
- Tyler, C. L., and M. Kowalewski. 2014. Utility of marine benthic associations as a multivariate proxy of paleobathymetry: A direct test from recent coastal ecosystems of North Carolina. *PloS One* **9**: e95711. doi:[10.1371/journal.pone.0095711](https://doi.org/10.1371/journal.pone.0095711)
- Tzeng, J., H. H.-S. Lu, and W.-H. Li. 2008. Multidimensional scaling for large genomic data sets. *BMC Bioinform.* **9**: 179. doi:[10.1186/1471-2105-9-179](https://doi.org/10.1186/1471-2105-9-179)
- Urbach, E., K. L. Vergin, G. L. Larson, and S. J. Giovannoni. 2007. Bacterioplankton communities of Crater Lake, OR: dynamic changes with euphotic zone food web structure and stable deep water populations. *Hydrobiologia* **574**: 161–177. doi:[10.1007/s10750-006-0351-5](https://doi.org/10.1007/s10750-006-0351-5)
- Vizzini, S., and A. Mazzola. 2004. Stable isotope evidence for the environmental impact of a land-based fish farm in the western Mediterranean. *Mar. Pollut. Bull.* **49**: 61–70. doi:[10.1016/j.marpolbul.2004.01.008](https://doi.org/10.1016/j.marpolbul.2004.01.008)
- White, J., M. Richard, M. Massot, and S. Meylan. 2011. Cloacal bacterial diversity increases with multiple mates: Evidence of sexual transmission in female common lizards. *PloS One*. **6**: e22339. doi:[10.1371/journal.pone.0022339](https://doi.org/10.1371/journal.pone.0022339)
- Wnuk, E., and A. Majid. 2014. Revisiting the limits of language: The odor lexicon of Maniq. *Cognition* **131**: 125–138. doi:[10.1016/j.cognition.2013.12.008](https://doi.org/10.1016/j.cognition.2013.12.008)
- Zhu, C., and J. Yu. 2009. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* **182**: 875–888. doi:[10.1534/genetics.108.098863](https://doi.org/10.1534/genetics.108.098863)
- Zuur, A., E. N. Ieno, and G. M. Smith. 2007. *Analyzing ecological data*. Springer.

Acknowledgments

Funding for this research was provided by a U.S. Environmental Protection Agency STAR grant (#FP91780901-0) awarded to E. Dexter and S. Bollens. Space used during manuscript preparation was generously provided to S. Bollens and G. Rollwagen-Bollens by the Oceans Institute and the School of Civil, Environmental and Mining Engineering, University of Western Australia.

Conflict of Interest

None declared.

Submitted 4 December 2017

Revised 26 March 2018

Accepted 13 May 2018

Associate editor: Malinda Sutor