



# Functional rarefaction: estimating functional diversity from field data

Steven C. Walker, Mark S. Poos and Donald A. Jackson

*S. C. Walker (steve.walker@utoronto.ca), M. S. Poos and D. A. Jackson, Dept of Ecology and Evolutionary Biology, 25 Harbord Street, Univ. of Toronto, Toronto, Ontario, Canada, M5S 3G5.*

Studies in biodiversity-ecosystem function and conservation biology have led to the development of diversity indices that take species' functional differences into account. We identify two broad classes of indices: those that monotonically increase with species richness (MSR indices) and those that weight the contribution of each species by abundance or occurrence (weighted indices). We argue that weighted indices are easier to estimate without bias but tend to ignore information provided by rare species. Conversely, MSR indices fully incorporate information provided by rare species but are nearly always underestimated when communities are not exhaustively surveyed. This is because of the well-studied fact that additional sampling of a community may reveal previously undiscovered species. We use the rarefaction technique from species richness studies to address sample-size-induced bias when estimating functional diversity indices. Rarefaction transforms any given MSR index into a family of unbiased weighted indices, each with a different level of sensitivity to rare species. Thus rarefaction simultaneously solves the problem of bias and the problem of sensitivity to rare species. We present formulae and algorithms for conducting a functional rarefaction analysis of the two most widely cited MSR indices: functional attribute diversity (FAD) and Petchev and Gaston's functional diversity (FD). These formulae also demonstrate a relationship between three seemingly unrelated functional diversity indices: FAD, FD and Rao's quadratic entropy. Statistical theory is also provided in order to prove that all desirable statistical properties of species richness rarefaction are preserved for functional rarefaction.

Walker et al. (1999) and Petchev and Gaston (2002) rekindled interest in developing indices of diversity that take into account ecologically relevant functional differences between species (Mason et al. 2003, 2005, Petchev et al. 2004, Ricotta 2004, Mouillot et al. 2005, Botta-Dukat 2005, Petchev and Gaston 2006, 2007, Podani and Schmera 2006). These functional diversity indices are motivated by the intuitive notion that, given constant species richness across a group of communities, the community with the most functionally similar species should be considered the least diverse of the group.

Not surprisingly, many different functional diversity indices have been proposed. They can be roughly divided into two types. Indices of the first type are characterized by the following intuitive property: they do not increase if a species is removed from the community and do not decrease if a species is added to the community (Petchev and Gaston 2006). Diversity indices with this property are said to be monotonically increasing with species richness (hereafter referred to as MSR indices). Examples of MSR indices include functional group richness, functional attribute diversity (FAD, Walker et al. 1999) and functional diversity (FD, Petchev and Gaston 2002). Indices of the second type are weighted by abundance or occurrence (hereafter referred to as weighted indices). Weighted indices

tend to have a weaker correlation with species richness than MSR indices. Examples of weighted indices of species diversity are Simpson's and Shannon's index. Examples of weighted functional diversity indices include Rao's (1982) quadratic entropy,  $FD_{VAR}$  (Mason et al. 2003) and Ricotta's (2004) expected taxonomic distinctiveness.

We highlight two major differences between MSR and weighted indices that are central to our paper. First, weighted indices are relatively insensitive to rare species whereas MSR indices are sensitive to rare and common species. Therefore, weighted indices may be more appropriate when rare species tend to be functionally unimportant. However, there may be many ecosystem functions that are highly sensitive to rare species. Power et al. (1996) provided for an excellent review of such ecosystem functions. For example, experimental evidence suggests that the removal of rare species can increase the susceptibility of grassland communities to invasions (Lyons and Schwartz 2001). Furthermore, rare species can be important in maintaining ecosystem function through periods of environmental change (Walker et al. 1999). In summation, weighted and MSR indices are likely to be complementary to one another, not competitive.

Second, weighted indices generally will be less biased than MSR indices. Because MSR indices are, by definition,

monotonic with species richness, they will always be underestimated if sampling effort is low. The reason for this is identical to the reason why observed levels of species richness underestimate community-level species richness when sampling effort is low; there may be some species to be discovered in the portions of the community that have not yet been sampled. We refer to this problem as a sample-size-induced bias. This problem of sample-size-induced bias is not likely to be an issue in highly controlled experiments of small spatial extent (e.g. Cedar Creek experiments, Tilman et al. 1997). This result is because small areas are easily surveyed exhaustively and the species composition itself has been manipulated and therefore need not be estimated. Recently however, there has been increasing interest in investigating the implications of these small-scale experimental results for the relationships between biodiversity, ecosystem functioning and climate at larger scales and in more naturally assembled communities (Loreau et al. 2001, 2003, Naeem and Wright 2003, Symstad et al. 2003, Cardinale et al. 2004, Tylianakis et al. 2006). However, typically it is not possible to exhaustively sample communities of large spatial extent. Weighted indices have the advantage of being easy to estimate in a relatively unbiased manner but largely ignore information provided by rare species; on the other hand, MSR indices take full advantage of the information provided by rare species but are difficult to estimate in an unbiased fashion.

In our paper, we show how the rarefaction technique (Sanders 1968, Hurlbert 1971, Smith and Grassle 1977, Colwell et al. 2004) from species richness studies can simultaneously address the problem of bias in MSR indices and unify MSR and weighted indices into a framework familiar to community ecologists studying species diversity. This technique allows researchers to make unbiased estimates of diversity that take full advantage of the statistical information provided by rare species. Specifically, we make the following contributions. First, we define rarefaction in a general sense so that it can be applied to functional diversity indices; this is necessary because rarefaction has been used almost exclusively to measure species and taxonomic diversity. This development allows us to prove that all desirable statistical properties of species richness rarefaction are preserved for functional rarefaction. Second, we review two well-known MSR indices, FAD and FD, and the two most common sampling models used in species richness rarefaction. Third, we present formulae and algorithms for conducting a functional rarefaction analysis with these indices and sampling models. This allows us to show that weighted indices arise naturally from the rarefaction of an MSR index. Specifically, we show that Rao's (1982) quadratic entropy arises from both FAD and FD rarefaction. Finally, as an example, we use functional rarefaction to analyze changes in the functional diversity of birds at the Hubbard Brook forest. Through the example, we also show how to use dimension-reduction techniques to check if functional rarefaction analyses are misleading.

## The conceptual framework of generalized rarefaction

Previously, rarefaction has only been applied to species richness,  $S$ . Therefore, it is necessary to generalize the

notion of rarefaction before we can apply it to functional diversity indices. Generalized rarefaction provides a framework for developing rarefied versions of other functional diversity indices not covered here and allows us to prove that the statistical properties of species richness rarefaction are preserved for functional rarefaction. Readers who are familiar with the rarefaction approach and who are not concerned with the theory of why rarefaction works in general may skip this section without loss of continuity.

We consider a generalized MSR diversity index,  $D$ , and a generalized sampling model which involves selecting (randomly or otherwise)  $v$  sampling units (hereafter referred to as units) from a community. The diversity index,  $D$ , represents the true unknown diversity of the entire community. We wish to estimate  $D$  by taking  $v$  sampling units from the community. In typical statistical settings we would use the information in the sample to make inferences about the entire community. Unfortunately, it is very difficult to use the diversity of a sample,  $\hat{D}$ , to make inferences about  $D$  because the species list in the sample will typically be shorter than the species list in the community. Therefore, the estimation of  $D$  must be made by extrapolation. Extrapolation, while necessary for estimating the diversity of entire communities of large extent, can be as dangerous in diversity estimation as it is in linear regression (Mao and Colwell 2005).

If one is more interested in comparing diversities across communities than in estimating the diversity of entire communities, the rarefaction technique (sometimes called interpolation) is often more accurate than extrapolation techniques (Colwell et al. 2004). The main idea behind rarefaction is to shift interest from the diversity of the entire community,  $D$ , to the expected diversity,  $\langle D_\mu \rangle$ , in  $\mu$  units. This expected diversity,  $\langle D_\mu \rangle$ , is referred to as the true rarefied diversity and  $\mu$  as the rarefied sample size. Like  $D$ ,  $\langle D_\mu \rangle$  is an unknown community-level quantity. However, unlike  $D$ , one can determine an unbiased estimator of  $\langle D_\mu \rangle$  if  $\mu$  is no greater than  $v$  (which is the size of the sample actually taken from the community). If we assume that  $v \geq \mu$ , and that the  $v$  units were randomly sampled in an independent manner, an unbiased estimator of  $\langle D_\mu \rangle$  is the expected diversity,  $\langle \hat{D}_\mu \rangle$ , in a sub-sample of  $\mu$  units without-replacement from the original sample of  $v$  units (see Appendix A for a proof). This result follows from an appropriate generalization of the work of Smith and Grassle (1977). The restriction of random and independent sampling can be removed (Appendix A). This expected diversity,  $\langle \hat{D}_\mu \rangle$ , is referred to as the sample rarefied diversity. Furthermore, for all specific cases considered in the following sections,  $\langle \hat{D}_\mu \rangle$  is the best unbiased estimator of  $\langle D_\mu \rangle$  in the sense that it has the smallest sampling variance of all unbiased estimators (see Appendix A for a proof).

## Functional diversity indices and sampling models

In the previous section we considered a generalized diversity index and sampling model. In applied settings, we need to define specific indices and sampling models in order to derive practical formulae and algorithms.

### Functional diversity indices

We apply generalized rarefaction to two well-known MSR functional diversity indices: functional attribute diversity, FAD, of Walker et al. (1999) and functional diversity, FD, of Petchey and Gaston (2002). These indices are both based on the idea of functional trait space. Consider a community composed of  $S$  species that are characterized by  $T$  functional traits. Let the value for trait  $t=1, \dots, T$  for species  $i=1, \dots, S$ , be  $X_{it}$ . The  $S$ -by- $T$  functional trait matrix,  $\mathbf{X}$ , has  $X_{it}$  as elements. The  $i$ th row of  $\mathbf{X}$  specifies the point in functional trait space that characterizes species  $i$ .

The FAD of a community is the total of all pair-wise distances between species in functional trait space. Let  $d_{ij}$  be the distance between the points that characterize species  $i$  and  $j$  in functional trait space. For simplicity we take  $d_{ij}$  to be the Euclidean distance in this paper but many other choices are possible (see for example Legendre and Legendre 1998). FAD is simply given by the sum of the pair-wise distances between species (Walker et al. 1999),

$$\text{FAD} = \sum_{i>j} d_{ij}$$

Petchey and Gaston's (2002) FD cannot be written compactly in terms of the pair-wise distances between species. FD is the total length of all of the branches of a dendrogram derived from the functional trait matrix,  $\mathbf{X}$ . Before FD can be calculated, a distance measure and clustering algorithm must be chosen; different choices give different results (Podani and Schmera 2006). Here we use Euclidean distances and an unweighted pair group method with arithmetic mean (UPGMA) to calculate branch lengths.

Petchey and Gaston (2006) emphasized the following difference between FD and FAD. Consider a species,  $A_S$ , that is functionally identical to a species that is a member of a particular community,  $A_C$ . If species  $A_S$  is added to community  $A_C$ , the FD of  $A_C$  is unchanged. Furthermore, if  $A_S$  is very similar, but not identical, to one or more species in  $A_C$ , FD will only increase a small amount if  $A_S$  is added to  $A_C$  whereas FAD will increase a larger amount. These properties of FD and FAD imply that FD is insensitive to redundant species whereas FAD is sensitive to redundant species. In our opinion, the sensitivity or insensitivity to redundant species cannot be used to argue for the general use of a particular index without considering the context of the study. The goals of the study and the nature of the study system need to be taken into account when deciding which functional diversity indices to use. For example, redundant species can help buffer the impacts of species loss on ecosystem function (Walker et al. 1999). Therefore, FAD might be a more appropriate index of the resilience of a community in the face of environmental change. However, redundant species are less likely to be important for maintaining ecosystem function in the short term. Therefore, FD might be more appropriate for measuring the aspects of functional diversity that have the potential to immediately impact ecosystem functioning. Another difference between FD and FAD is that FD requires the subjective choice of a clustering algorithm in addition to a distance measure whereas FAD only requires the choice of a distance measure (Petchey and Gaston 2006). Our purpose here is not to argue in favor of either

index, but rather to develop sound methods of inference for each. With the inferential tools that we develop here, researchers will be better able to evaluate the properties of each measure in the light of field data.

### Sampling models

Following Gotelli and Colwell (2001), we consider two very common forms of data: abundance data and sample data. In doing so, the abstract concept of sampling effort is made more practical by measuring it in units of individuals (for abundance data) or plots (for sample data). Each type of data is modeled using a multinomial distribution.

For abundance data, consider a random sample of  $N$  individuals from a community containing  $S$  species. Each individual is identified to species. This model is a special case of our generalized sampling model. Sampling effort is measured in units of individuals ( $v=N$ ). We wish to estimate the expected diversity,  $\langle D_n \rangle$ , in  $n \leq N$  individuals ( $\mu=n$ ). We estimate  $\langle D_n \rangle$  with the expected diversity,  $\langle \hat{D}_n \rangle$ , in a sub-sample, without replacement, of  $n$  individuals from the original sample of  $N$  individuals.

For sample data, consider a random sample of  $M$  plots (or quadrats, mist nets, pheromone traps, etc.) from a community containing  $S$  species. Because the community contains  $S$  species, each plot could be characterized by one of the  $2^S$  possible species lists (or  $2^S - 1$  species lists if the sampling method ensures that at least one species is sampled in every plot). For convenience, we give an index,  $k=1, \dots, 2^S$ , to each of the possible species lists. It does not matter how the species lists are numbered as long as we are consistent. This model is also a special case of our generalized sampling model. Sampling effort is measured in units of plots ( $v=M$ ). We wish to estimate the expected diversity,  $\langle D_m \rangle$ , in  $m \leq M$  plots ( $\mu=m$ ). We estimate  $\langle D_m \rangle$  with the expected diversity,  $\langle \hat{D}_m \rangle$ , in a sub-sample, without replacement, of  $m$  plots from the original sample of  $M$  plots.

The sample-based model is, in a sense, more informative than the abundance-based model. This is because by allowing the possibility of sampling more than one individual at a time, the data from the sample-based model contains information on the spatial co-occurrence of species at the grain of the plot (Colwell et al. 2004). In contrast, data from the abundance-based model only provides information on the abundances of the species.

### Functional rarefaction

We apply generalized rarefaction to the measurement of functional diversity. All of the results presented in this section follow directly from the generalized rarefaction framework, the two sampling models and the definitions of FAD and FD.

### Formulae and algorithms

We describe the formulae and algorithms that are required to conduct a functional rarefaction analysis of field data. Comparisons with the more familiar species richness rarefaction are made to further illustrate how functional rarefaction is related to species richness rarefaction; however,

the formulae for species richness rarefaction were originally derived by Hurlbert (1971), Uglund et al. (2003) and Colwell et al. (2004). Proofs and derivations of formulae are given in Appendices A and B. MATLAB code is available from the first author upon request.

We have organized our results in Table 1 where its organization reflects the relationships between the types of rarefaction analysis and their relationship with the generalized rarefaction framework. Table 1 organizes 12 computational recipes (eight formulae and four algorithms). Each of these 12 recipes is categorized in three ways: by diversity index ( $D=S$ ,  $FAD$  or  $FD$ ), by sampling model (abundance-based or sample-based) and by statistical setting (unknown true rarefied diversity or known sample rarefied diversity). Because rarefaction is most useful when we have not sampled the entire community, the formulae and algorithms for true rarefied diversities will not be used in most applied settings. The definitions of the symbols used in Table 1 are presented in Table 2.

We were unable to find useful formulae for calculating rarefied  $FD$  indices. Instead we describe simple Monte Carlo algorithms for approximating these quantities (Table 3). Algorithms A and C amount to simulating the sampling process for  $\langle FD_n \rangle$  and for  $\langle FD_m \rangle$  from the abundance-based model and the incidence-based model respectively. Algorithms B and D amount to sub-sampling individuals and plots respectively, without replacement, from the original sample. The Monte Carlo sample average for each of these algorithms is approximately equal to the  $FD$  index of interest. The approximations improve as the number of Monte Carlo samples increases.

We use a simple bootstrap to calculate confidence intervals for each of the six types of rarefaction estimates in Table 1 (see Appendix C for details). Researchers likely will be more interested however in examining how rarefied functional diversity indices are related to a covariate of interest rather than simply putting confidence intervals on the points of a rarefaction curve. For example, we may wish to ask: is  $\langle FD_n \rangle$  linearly related to an environmental variable that characterizes each of the communities? In Appendix C, we describe a simple bootstrap to calculate the confidence interval of the slope of a least-squares linear regression of a rarefied diversity index on a covariate of interest. We use this procedure in our Hubbard Brook bird community example. Note that this bootstrap procedure can be used for most regression models.

### Unifying MSR and weighted functional diversity indices

It is known that the abundance-based rarefaction of species richness yields a family of weighted indices. For example, when species richness is rarefied, Simpson's (1949) index is recovered as  $\langle S_2 \rangle - 1$  (Smith and Grassle 1977). As the rarefied sample size,  $n$ , increases from 2, weighted indices are obtained that are progressively more sensitive to rare species. As  $n$  further increases to infinity (or to the total number of individuals in the community), species richness is obtained. Thus rarefaction unifies species richness and weighted species diversity indices such as Simpson's (1949) index.

The effect that the rarefied sample size has on the sensitivity to rare species is preserved in generalized rarefaction. For example, consider Rao's (1982) quadratic entropy,

Table 1. Summary of computational formulae and algorithms for rarefaction analysis.

	D = S	D = FAD	D = FD
Abundance-based rarefaction True rarefied diversity, $\langle D_n \rangle$	$\langle S_n \rangle = S - \sum_{i=1}^S (1 - p_i)^n$	$\langle FAD_n \rangle = FAD - \sum_{i>j} d_{ij} [(1 - p_i)^n + (1 - p_j)^n - (1 - p_i - p_j)^n]$	$\langle FD_n \rangle \rightarrow$ algorithm A
Sample rarefied diversity, $\langle \hat{D}_n \rangle$	$\langle \hat{S}_n \rangle = \hat{S} - \sum_{i=1}^{\hat{S}} \frac{\binom{N - N_i}{n}}{\binom{N}{n}}$	$\langle \hat{FAD}_n \rangle = \hat{FAD} - \frac{1}{\binom{N}{n}} \sum_{i>j} d_{ij} \left[ \binom{N - N_i}{n} + \binom{N - N_j}{n} - \binom{N - N_i - N_j}{n} \right]$	$\langle \hat{FD}_n \rangle \rightarrow$ algorithm B
Sample-based rarefaction True rarefied diversity, $\langle D_m \rangle$	$\langle S_m \rangle = S - \sum_{i=1}^S (1 - q_i)^m$	$\langle FAD_m \rangle = FAD - \sum_{i>j} d_{ij} [(1 - q_i)^m + (1 - q_j)^m - (1 - q_i - q_j)^m]$	$\langle FD_m \rangle \rightarrow$ algorithm C
Sample rarefied diversity, $\langle \hat{D}_m \rangle$	$\langle \hat{S}_m \rangle = \hat{S} - \sum_{i=1}^{\hat{S}} \frac{\binom{M - Y_i}{m}}{\binom{M}{m}}$	$\langle \hat{FAD}_m \rangle = \hat{FAD} - \frac{1}{\binom{M}{m}} \sum_{i>j} d_{ij} \left[ \binom{M - Y_i}{m} + \binom{M - Y_j}{m} - \binom{M - Y_{ij}}{m} \right]$	$\langle \hat{FD}_m \rangle \rightarrow$ algorithm D

The table is organized in three ways: 1) individual- versus sample-based, 2) community- versus sample-level and 3) diversity index ( $S$ ,  $FAD$  or  $FD$ ). All symbols in the table are defined in the main text and in Table 2. The individual-based formulae for species richness were originally derived by Hurlbert (1971). The sample-based formulae for species richness were originally derived independently by Uglund et al. (2003) and Colwell et al. (2004).

Table 2. Definitions of variables related to the rarefied diversity indices in Table 1 and Appendix A.

Type of variable	Variable	Description
Non-rarefied diversity indices (see Table 1 for rarefied indices)	D	a generalized diversity index for the entire community
	$\hat{D}$	a generalized diversity index for the sample
	FAD	functional attribute diversity for the entire community
	$\hat{FAD}$	functional attribute diversity for the sample
	FD	Petchey and Gaston's functional diversity for the entire community
	$\hat{FD}$	Petchey and Gaston's functional diversity for the sample
	Q	Rao's quadratic entropy for the entire community
	$\hat{Q}$	Rao's quadratic entropy for the sample
Functional trait information	S	total number of species in the entire community
	$\hat{S}$	total number of species in the sample
	$d_{ij}$	distance between species i and j in functional trait space
	T	total number of functional traits
Abundance-based sampling model	$X_{it}$	value of functional trait t for species i
	<b>X</b>	functional trait matrix (with the $X_{it}$ as elements)
	n	rarefied sample size in units of individuals
Sample-based sampling model	N	sample size in units of individuals
	$N_i$	observed number of individuals in species i
	$p_i$	probability that a randomly selected individual belongs to species i
Generalized sampling model	m	rarefied sample size in units of plots
	M	sample size in units of plots
	$M_k$	observed number of plots with species-list k (species-list abundance)
	$q_i$	probability that a randomly selected plot contains species i
	$q_{ij}$	probability that a randomly selected plot contains species i,j or both
	$r_k$	probability that a randomly selected plot contains species-list k
	$Y_i$	observed number of plots with species i
	$Y_{ij}$	observed number of plots with species i,j or both
	<b><math>\theta</math></b>	the parameters for the generalized sampling model
	$\mu$	rarefied sample size under the generalized sampling model
	$\nu$	sample size under the generalized sampling model
	<b>v</b>	a sample from the generalized sampling model

Q, a weighted index. Q is the expected distance in functional trait space between two randomly selected individuals from the community. It follows that Q is equal to  $\langle FAD_2 \rangle$  for abundance data. We also have that  $2Q$  is equal to  $\langle FD_2 \rangle$ . This follows because there must be at least be three species for clustering to occur in a dendrogram; hence, the FD of a two species community is simply equal to two times the distance

between the two species in functional trait space. This result demonstrates a strong connection between three seemingly unrelated functional diversity indices, FAD, FD and Q. Weighted indices also arise naturally from sample-based rarefaction. Thus, functional rarefaction unifies MSR and weighted indices in a manner that takes advantage of the strengths of each type of index.

Table 3. Algorithms for calculating the four rarefied forms of FD.

Algorithm A: approximating $\langle FD_n \rangle$	
1	Generate a species list for a random sample of n individuals based on the species probabilities, $p_i$ , $i = 1, \dots, S$ .
2	Calculate the FD for this species list.
3	Repeat many times.
4	$\langle FD_n \rangle \approx$ the average of the randomly generated FD.
Algorithm B: approximating $\langle \hat{FD}_n \rangle$	
1	Generate a species list for a random sample of n individuals, without replacement, from the individuals that were sampled from the community.
2	Calculate the FD for this species list.
3	Repeat many times.
4	$\langle \hat{FD}_n \rangle \approx$ the average of the randomly generated FD.
Algorithm C: approximating $\langle FD_m \rangle$	
1	Generate a species list for a random sample of m plots based on the species list probabilities, $r_k$ , $k = 1, \dots, 2^S$ .
2	Calculate the FD for this species list.
3	Repeat many times.
4	$\langle FD_m \rangle \approx$ the average of the randomly generated FD.
Algorithm D: approximating $\langle \hat{FD}_m \rangle$	
1	Generate a species list for a random sample of m plots, without replacement, from the plots that were sampled from the community.
2	Calculate the FD for this species list.
3	Repeat many times.
4	$\langle \hat{FD}_m \rangle \approx$ the average of the randomly generated FD.

## Example: Hubbard Brook bird community

To exemplify how rarefaction can be used to correct for sample-size differences in the comparison of communities based on functional diversity, we analyze a bird community data set from Hubbard Brook forest (Holmes and Sturges 1975, Holmes et al. 1979). As this is a very large data set, we use only the species abundances from samples collected in late spring between 25 June and 9 July for the years 1969 to 1973 (Holmes and Sturges 1975). Holmes et al. (1979) provided a functional trait matrix with 27 foraging behavior traits. This matrix was used by Petchey and Gaston (2002) to exemplify the use of FD as a functional diversity index.

As an example research question we ask: does late spring bird diversity (as measured by  $S$ , FAD and FD) decline between 1969 and 1973 in the Hubbard Brook forest? Because the number of individuals surveyed varied from year to year (range = 234 to 355), abundance-based rarefaction is required before the relationship between year and functional (or species) diversity can be studied. We estimate  $\langle S_n \rangle$ ,  $\langle FAD_n \rangle$  and  $\langle FD_n \rangle$  for  $n = 2$  and 150. Given  $\langle FD_2 \rangle = 2\langle FAD_2 \rangle$ , we compute  $\langle FAD_2 \rangle = Q$ , where  $Q$  is Rao's (1982) quadratic entropy, and not  $\langle FD_2 \rangle$ . This results in five diversity indices. We did not choose the rarefied sample size,  $n$ , to be too close to the sample size of the year with the smallest sample size (1973,  $n = 234$ ) because the bootstrap sampling distributions were skewed for those values of  $n$  for some communities (see Appendix C for a discussion of the importance of symmetric bootstrap distributions). We compute bootstrap confidence intervals for each of the five indices for each year and bootstrap confidence intervals for the least-squares slope of the linear regression model with year as the independent variable (Appendix C).

We now provide a rationale for choosing these five indices and a framework with which to interpret ecologically the results. For this purpose, we characterize each of the five diversity indices using three properties. First,  $\langle FAD_2 \rangle$ ,  $\langle FAD_{150} \rangle$  and  $\langle FD_{150} \rangle$  are sensitive to the magnitude of species differences in trait-space whereas  $\langle S_2 \rangle$  and  $\langle S_{150} \rangle$  are not. Second,  $\langle S_{150} \rangle$ ,  $\langle FAD_{150} \rangle$  and  $\langle FD_{150} \rangle$  are relatively sensitive to rare species whereas  $\langle S_2 \rangle$  and  $\langle FAD_2 \rangle$  are not. Third,  $\langle FAD_2 \rangle$ ,  $\langle FAD_{150} \rangle$ ,  $\langle S_2 \rangle$  and  $\langle S_{150} \rangle$  are relatively sensitive to functionally redundant species whereas  $\langle FD_{150} \rangle$  is not. Therefore, these five indices collectively should provide reasonable sensitivity to several important aspects of functional and species diversity.

The slope of the least-squares regression line for all five indices was negative (Fig. 1) indicating a decline in diversity, no matter how it is measured. However, only three of the slopes were significantly different from zero:  $\langle S_2 \rangle$ ,  $\langle S_{150} \rangle$  and  $\langle FAD_{150} \rangle$ . These results suggest a fairly clear ecological interpretation. Late spring bird diversity was most likely declining in the Hubbard Brook forest between 1969 and 1973. This is evidenced by the estimated negative slope for all five diversity indices regressed against year. Because we are using more than one type of diversity index, we can make more detailed conclusions. Two results lead to the conclusion that it is largely the functionally redundant species that are being lost over time. First, the only index that is insensitive to redundant species,  $\langle FD_{150} \rangle$ , did not show a significant decline. Second, three of the four indices

that are sensitive to redundant species did show a significant decline.

Despite the fact that some common species are becoming less common over time ( $\langle S_2 \rangle$  is significantly declining), we did not detect a significant decline in the functional differences between common species (as measured by  $\langle FAD_2 \rangle$ ). This suggests that the common species that become less common over time are not sufficiently unique with respect to their functional traits to result in a significant decline in  $\langle FAD_2 \rangle$ . However, overall our results suggest that late spring bird diversity was declining, in the Hubbard Brook forest between 1969 and 1973, largely because of losses of, and declines in, redundant species.

While summary statistics, such as rarefied diversity indices, are invaluable tools for describing and making inferences about complex data, we should always check that they are not misleading. To verify our conclusion that observed diversity declines largely result from losses in redundant species, we look at the data more closely. Since, the estimated slope of the  $\langle S_{150} \rangle$  regression is approximately  $-0.57$ , this indicates that roughly two species have been lost from samples of 150 individuals over the study period. Therefore, if our conclusions are correct, the two least abundant species in 1973 should be close, in some sense, to other species in functional trait space; this would indicate the species being lost are redundant in terms of the 27 functional traits measured. Because the functional trait space has 27 dimensions, it can not be easily represented on graph paper. We therefore used principal coordinates analysis (PCoA) with the Euclidean distance measure to reduce the dimensionality of the trait space to two axes (Gower 1966); note that a PCoA with Euclidean distances is equivalent to a principal components analysis (PCA) on the observed covariance matrix (Jackson 1993). The Euclidean distance is appropriate since we used this measure to calculate FAD and FD. The first two PCoA axes account for 67% of the variation between species in functional trait space. The 22 bird species are plotted on this reduced functional trait space (Fig. 2). This plot clearly shows that the two least abundant species in 1973 are quite similar to other species that have persisted over the study period. Therefore, the conclusions of the functional rarefaction analysis appear to be appropriate.

## Discussion

We have demonstrated how to use the rarefaction technique to correct for sample-size-induced biases in the estimation of two functional diversity indices, FAD and FD. For the first time, we now have a method for comparing these functional diversity indices between natural communities of large spatial extent, in a manner that takes advantage of the statistical information in rare species, in the common situation when there is variable sampling effort between communities. This opens new possibilities for researchers as these types of studies are likely to become more frequent as interest grows in questions about how the insights gained through the plot-scale biodiversity-ecosystem function experiments of the last 10 to 15 years will scale up to larger spatial extents (Loreau et al. 2001, Naeem and Wright 2003, Symstad et al. 2003).

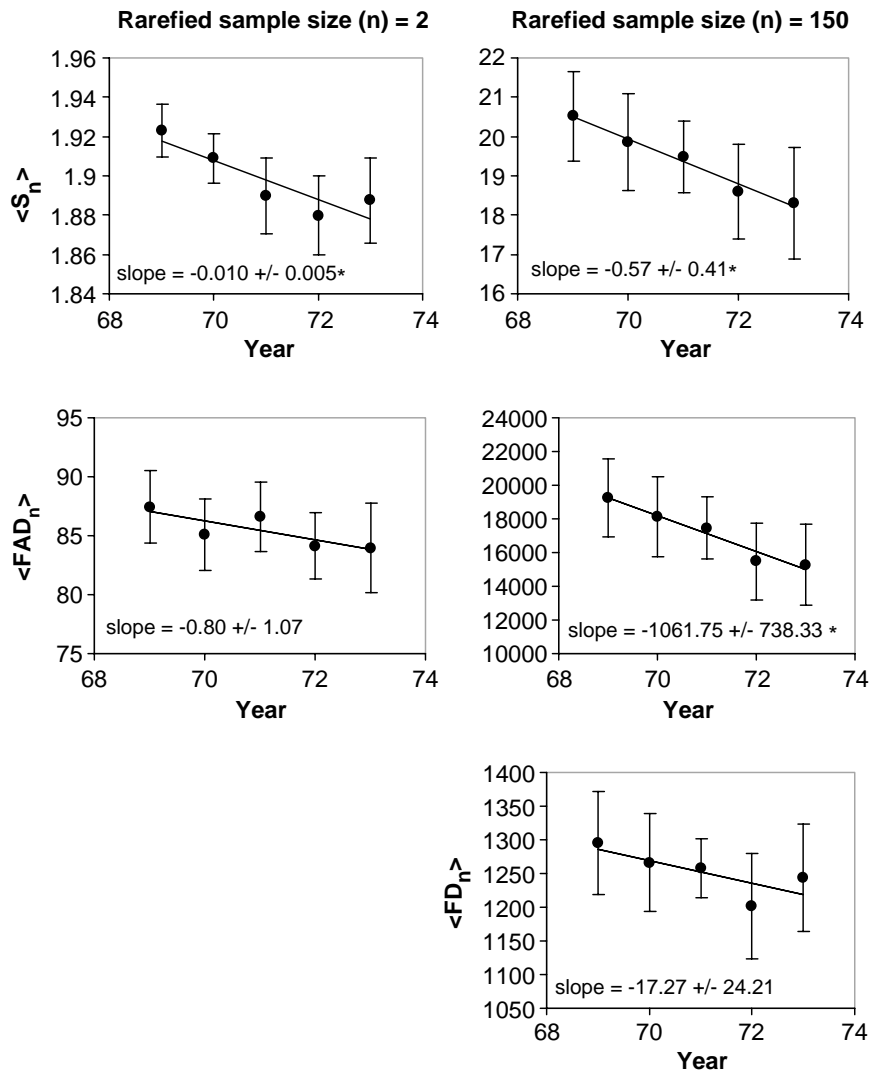


Fig. 1. Changes in the functional and species diversity of the early spring bird community in Hubbard Brook forest between 1969 and 1973. Five diversity indices are used:  $\langle S_2 \rangle$ ,  $\langle FAD_2 \rangle$ ,  $\langle S_{150} \rangle$ ,  $\langle FAD_{150} \rangle$  and  $\langle FD_{150} \rangle$ . Dots and error bars are estimated levels of diversity with 95% bootstrap confidence intervals. Lines are least-squares lines. The slopes of these lines are given along with 95% bootstrap confidence intervals. An asterisk indicates significance in the sense that the confidence interval does not overlap a slope of zero.

Much of the previous work on functional diversity indices focused on the relative strengths and weaknesses of the various indices (Naeem and Wright 2003, Petchey and Gaston 2006, 2007, Podani and Schmera 2006). We have intentionally avoided picking one side or the other in this debate, but rather focused more on developing methods of statistical inference about the various indices. By avoiding this debate, we were also able to demonstrate three practical techniques of statistical analysis. First, we showed how to investigate the relationship between rarefied functional diversity indices (in this case for a bird community) and a covariate of interest (in this case time). Walker et al. (1999) studied the relationship between FAD and distance from a water source in Australian rangelands. However, they were unable to make any statistical inferences from these data. Our functional rarefaction methods make such inferences possible. Second, we showed via the bird community example, how various indices are complementary; each of the five indices that we used quantified different aspects of community structure. Third, we used PCoA to visualize

functional trait data for the purpose of corroborating conclusions drawn from an analysis of functional diversity indices.

In addition to this practical approach, we provided a theoretical basis for why rarefaction works in general. To do this, we developed a mathematical framework of generalized rarefaction, which contains classical species richness rarefaction as a special case (Sanders 1968, Hurlbert 1971, Smith and Grassle 1977, Colwell et al. 2004). Using generalized rarefaction, we proved that the rarefaction of any diversity index shares the same statistical properties as species richness rarefaction; in particular generalized rarefaction always yields unbiased estimates of rarefied diversity indices. Furthermore, rarefaction estimators are also the best of all unbiased estimators in the sense that they have minimum sampling variance if certain rather mild technical conditions are met (Appendix A). The well-studied relationship between the sensitivity to rare species and the rarefied sample size is also preserved through the generalization of rarefaction to all MSR indices; in particular, indices become

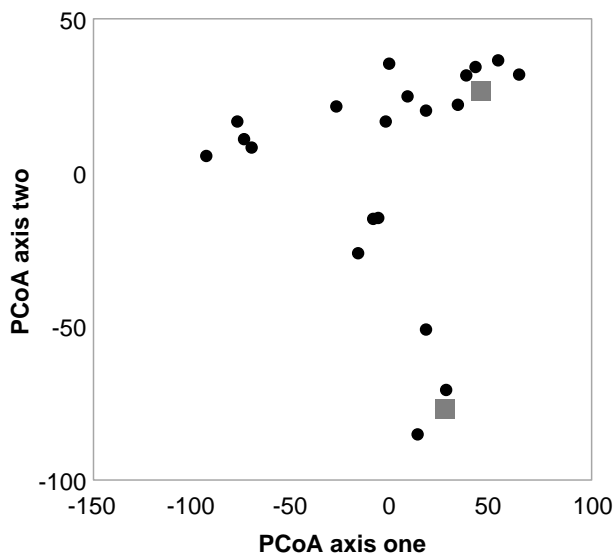


Fig. 2. Principal coordinates analysis (PCoA) of the functional trait matrix for the Hubbard Brook forest. The two least-abundant species in 1973 are squares, all other species are circles. Note how close these two rare species are to other species in functional trait space; this reinforces the conclusions drawn from functional rarefaction analysis that the diversity decline shown in Fig. 1 is largely due to losses in redundant species.

less sensitive to rare species as the rarefied sample size decreases. Therefore, rarefaction unifies MSR and weighted indices. For example, we showed that Rao's quadratic entropy,  $Q$ , an index that is relatively insensitive to rare species, is equivalent to a rarefied form of both FAD and FD.

Our work extends, generalizes and enhances previous work on bias and rarity in functional diversity indices. Bady et al. (2005) studied bias in Rao's (1982) quadratic entropy,  $Q$ . Calculating  $Q$  requires the relative abundances of the species in the community and the pair-wise distances in functional trait space between those species. To estimate  $Q$ , Bady et al. (2005) simply substituted the relative abundances of the species in the sample. They conjectured that this estimate is slightly biased but that the bias quickly disappears with more sampling effort. The basis for this conjecture was the observation that estimates of  $Q$  leveled off with sampling effort for the invertebrate community that they analyzed. In this paper, we prove that  $\langle \text{FAD}_2 \rangle$  (Table 1) is an unbiased estimate of  $Q$ , which is equivalent to  $\langle \text{FAD}_2 \rangle$ , for abundance-based rarefaction. In fact, our work highlights the benefits of using indices, such as  $Q = \langle \text{FAD}_2 \rangle$ , for which unbiased estimators exist. Ricotta (2004) developed a new family of functional diversity indices called expected taxonomic distinctiveness. This family of indices is a special case of the generalized rarefaction approach taken here. However, it differs from the present work in several important respects. First, a statistical treatment is not given. In particular, sampling variability and hypothesis testing are not considered and only the unknown community-level indices are discussed, which are biased if communities are not exhaustively sampled. Second, only one index is considered for rarefaction: the taxonomic distinctiveness. Third, no recommen-

dations or examples are given on how to interpret analyses based on expected taxonomic diversities.

Sampling processes in ecological field studies are often quite complex. For example, the Hubbard Brook data set that we use was collected in a manner that did not meet strictly the assumptions of our abundance-based sampling model. The dataset comprises all individuals living within a single ten hectare plot within the forest. Strictly speaking, this violates the model assumption of the random selection of individuals over the entire extent of the forest. As a second example, each point in trait space, that characterizes a species, is assumed to be a known constant by our abundance-based and sample-based models. This is unlikely to be true in most cases of interest. These types of violations have the potential to affect inferences. Future work should focus on solutions to these problems of model misspecification. For example, how robust are our methods to violations of model assumptions or can we develop more realistic sampling models?

When using FAD and FD, it is important to remember that the choice of distance measure, which quantifies the proximity of species in functional trait space, and the clustering algorithm, which prescribes how dendrograms are constructed for the calculation of FD, can influence the results of analyses (Podani and Schmera 2006). It is currently unclear how sensitive ecological conclusions are to these choices (Petchey and Gaston 2007). Understanding these issues remains an important research issue.

The statistical ecology of functional diversity is only just beginning to mature. This study provides functional diversity researchers with new statistical tools that allow the unbiased estimation and comparison of functional diversity indices, at scales larger than fully surveyed plots, when sampling effort may not be standardized across communities. We also showed how the statistical literature on species diversity can be used both as a model to anticipate statistical issues surrounding functional diversity (e.g. sample-size-induced bias) and as inspiration on how to solve new statistical issues in functional diversity as they arise (e.g. rarefaction).

*Acknowledgements* – We thank R. K. Colwell, N. Mason, S. Sharma and L. L. Timms for helpful comments on the manuscript. The manuscript benefited from discussions with J. Hughes. We also thank all of the researchers at the Hubbard Brook experimental forest who collected the bird community data used for our example. Funding for this research was provided by a Natural Sciences and Engineering Research Council of Canada (NSERC) Research Discovery Grant to DAJ, an Ontario Graduate Scholarship to MSP and an NSERC Graduate Scholarship to SCW.

## References

- Bady, P. et al. 2005. Use of invertebrate traits for the biomonitoring of European large rivers: the effects of sampling effort on genus richness and functional diversity. – *Freshwater Biol.* 50: 159–173.
- Botta-Dukat, Z. 2005. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. – *J. Veg. Sci.* 16: 533–540.



- Cardinale, B. J. et al. 2004. Effects of species diversity on the primary productivity of ecosystems: extending our spatial and temporal scales of inference. – *Oikos* 104: 437–450.
- Colwell, R. K. et al. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. – *Ecology* 85: 2717–2727.
- Gotelli, N. J. and Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. – *Ecol. Lett.* 4: 379–391.
- Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. – *Biometrika* 53: 325–338.
- Holmes, R. T. and Sturges, F. W. 1975. Bird community dynamics and energetics in a northern hardwoods ecosystem. – *J. Anim. Ecol.* 44: 175–200.
- Holmes, R. T. et al. 1979. Guild structure of the Hubbard Brook bird community: a multivariate approach. – *Ecology* 60: 512–520.
- Hurlbert, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. – *Ecology* 52: 577–586.
- Jackson, D. A. 1993. Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardizations, measures of association, and ordination methods. – *Hydrobiologia* 268: 9–26.
- Legendre, P. and Legendre, L. 1998. *Numerical ecology* (2nd English ed.). – Elsevier.
- Loreau, M. et al. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. – *Science* 294: 804–808.
- Loreau, M. et al. 2003. Biodiversity as spatial insurance in heterogeneous landscapes. – *Proc. Natl Acad. Sci. USA* 100: 12765–12770.
- Lyons, K. G. and Schwartz, M. W. 2001. Rare species loss alters ecosystem function — invasion resistance. – *Ecol. Lett.* 4: 358–365.
- Mao, C. X. and Colwell, R. K. 2005. Estimation of species richness: mixture models, the role of rare species and inferential challenges. – *Ecology* 86: 1143–1153.
- Mason, W. H. N. et al. 2003. An index of functional diversity. – *J. Veg. Sci.* 14: 571–587.
- Mason, N. W. H. et al. 2005. Functional richness, functional evenness and functional divergence: the primary components of functional diversity. – *Oikos* 111: 112–118.
- Mouillot, D. et al. 2005. Functional regularity: a neglected aspect of functional diversity. – *Oecologia* 142: 353–359.
- Naeem, S. and Wright, J. P. 2003. Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. – *Ecol. Lett.* 6: 567–579.
- Petchey, O. L. and Gaston, K. J. 2002. Functional diversity (FD), species richness and community composition. – *Ecol. Lett.* 5: 402–411.
- Petchey, O. L. and Gaston, K. J. 2006. Functional diversity: back to basics and looking forward. – *Ecol. Lett.* 9: 741–758.
- Petchey, O. L. and Gaston, K. J. 2007. Dendrograms and measuring functional diversity. – *Oikos* 116: 1422–1426.
- Petchey, O. L. et al. 2004. How do different measures of functional diversity perform. – *Ecology* 85: 847–857.
- Podani, J. and Schmera, D. 2006. On dendrogram-based measures of functional diversity. – *Oikos* 115: 179–185.
- Power, M. E. et al. 1996. Challenges in the quest for keystones. – *Bioscience* 46: 609–620.
- Rao, C. R. 1982. Diversity and dissimilarity coefficients – a unified approach. – *Theor. Popul. Biol.* 21: 24–43.
- Ricotta, C. 2004. A parametric diversity measure combining the relative abundances and taxonomic distinctiveness of species. – *Div. Distr.* 10: 143–146.
- Sanders, H. L. 1968. Marine benthic diversity: a comparative study. – *Am. Nat.* 102: 243–282.
- Simpson, E. H. 1949. Measurement of diversity. – *Nature* 163: 688.
- Smith, W. and Grassle, J. F. 1977. Sampling properties of a family of diversity measures. – *Biometrics* 33: 283–292.
- Symstad, A. J. et al. 2003. Long-term and large-scale perspectives on the relationships between biodiversity and ecosystem functioning. – *Bioscience* 53: 89–98.
- Tylianakis, J. M. et al. 2006. Diversity, ecosystem function, and stability of parasitoid-host interactions across a tropical habitat gradient. – *Ecology* 87: 3047–3057.
- Tilman, D. et al. 1997. Biodiversity and ecosystem properties. – *Science* 278: 1866–1867.
- Ugland, K. I. et al. 2003. The species-accumulation curve and estimation of species richness. – *J. Anim. Ecol.* 72: 888–897.
- Walker, B. H. et al. 1999. Plant attribute diversity, resilience, and ecosystem function: the nature and significance of dominant and minor species. – *Ecosystems* 2: 95–113.

## Appendix A. Properties of rarefaction estimators (generalizing Smith and Grassle 1977)

Here we show that  $\langle \hat{D}_\mu \rangle$  is an unbiased estimator for  $\langle D_\mu \rangle$ . We begin with some definitions. Let the estimator,  $\langle \hat{D}_\mu \rangle$ , be a random variable that depends on the sampling model indexed by a parameter,  $\theta$ . Let the sample of  $v$  units be denoted as  $\mathbf{v}$  and diversity in a sample of  $\mu < v$  units be  $\hat{D}_\mu$ . We define  $\langle \hat{D}_\mu \rangle \equiv E(\hat{D}_\mu | \mathbf{v})$ , where  $E(A | \mathbf{v})$  is the mathematical expectation of any random variable,  $A$ , given that we observed the data,  $\mathbf{v}$ . If  $\mathbf{v}$  arises from independent random sampling, as in the main text,  $E(A | \mathbf{v})$  is the expectation of  $A$  given that we randomly sub-sample from  $\mathbf{v}$  without-replacement. We also define  $\langle D_\mu \rangle \equiv E(\hat{D}_\mu | \theta)$  where  $E(A | \theta)$  is the mathematical expectation of  $A$  given that we sample from the model indexed by parameter  $\theta$ . These definitions of  $\langle \hat{D}_\mu \rangle$  and  $\langle D_\mu \rangle$  are more precise and more general than the definitions given in the main text. They also provide sufficient notation to prove that  $\langle \hat{D}_\mu \rangle$  is an unbiased estimator for  $\langle D_\mu \rangle$  in a succinct manner.

By the definition of an unbiased estimator, we are required to show that  $E(\langle \hat{D}_\mu \rangle | \theta) = \langle D_\mu \rangle$ . By substitution,  $E(\langle \hat{D}_\mu \rangle | \theta) = E(E(\hat{D}_\mu | \mathbf{v}) | \theta)$ . Because  $\theta$  is constant (but unknown), expectation that conditions on  $\theta$  is identical to unconditional expectation. Therefore, by the theorem of total expectation,  $E(\langle \hat{D}_\mu \rangle | \theta) = E(\hat{D}_\mu | \theta)$ . But, by definition,  $\langle D_\mu \rangle = E(\hat{D}_\mu | \theta)$  which proves the required equality,  $E(\langle \hat{D}_\mu \rangle | \theta) = \langle D_\mu \rangle$ .

The previous paragraph showed that for our general rarefaction model,  $\langle \hat{D}_\mu \rangle$  is an unbiased estimator of  $\langle D_\mu \rangle$ . However, lack of bias is only one desirable property of an estimator. Another desirable property is low sampling variance. These two desirable properties inspire statisticians to search for unbiased estimators that have the lowest possible sampling variance anywhere on the parameter space. Such an estimator, if it exists, is referred to as a uniformly minimum-variance, unbiased (UMVU) estimator. It follows directly from the Rao-Blackwell and Lehmann-Scheffé theorems that  $\langle \hat{D}_\mu \rangle$  is UMVU for  $\langle D_\mu \rangle$  if  $\mathbf{v}$  is a complete sufficient statistic. While this might not be true in general, it is true for the two ecological sampling models that we consider here because both are multinomial models. A complete sufficient statistic for any multinomial model is the vector of observed counts. For

our abundance-based model, this vector is  $\mathbf{v} = [N_1, \dots, N_S]$ , the vector of species abundances, and for our sample-based model it is  $\mathbf{v} = [M_1, \dots, M_{2s}]$ , the vector of species-list abundances. Our estimators for FD rarefaction are calculated by Monte Carlo. Therefore, the added Monte Carlo simulation error will increase the variance of the estimators above the UMVU case. However, we can make our Monte Carlo estimators arbitrarily close to the UMVU estimators by increasing the number of Monte Carlo iterations.

## Appendix B. FAD rarefaction formulae

In the following subsection we derive the four different types of FAD rarefaction,  $\langle \text{FAD}_n \rangle$ ,  $\langle \text{FAD}_m \rangle$ ,  $\langle \hat{\text{FAD}}_n \rangle$ ,  $\langle \hat{\text{FAD}}_m \rangle$ . One way to look at differences amongst these types of rarefaction is to recognize that they each arise from a different sampling model. The sampling models are categorized as abundance- or sample-based and as with- or without-replacement. Abundance- and sample-based rarefactions are indicated by an  $n$  and  $m$  respectively. Without-replacement rarefactions have a ‘hat’ over the FAD symbol and with-replacement rarefactions have no such ‘hat.’ These four models have crucially different biological and statistical interpretations. The with-replacement models are indexed by community-level parameters (the species and species-list probabilities) whereas the without-replacement models are indexed by sample-level parameters (the observed species and species-list abundances). Mathematically, however, these four types of rarefaction have very similar underpinnings and so we derive them all simultaneously. Let  $\langle \Phi_\mu \rangle$  be an FAD rarefaction for rarefied sample size,  $\mu$ ; that is,  $\langle \Phi_\mu \rangle$  can take values on the set,  $\{\langle \text{FAD}_n \rangle, \langle \text{FAD}_m \rangle, \langle \hat{\text{FAD}}_n \rangle, \langle \hat{\text{FAD}}_m \rangle\}$ .

We now provide a precise definition of rarefied FAD. We begin by defining some sets. Let  $\sigma = \{1, \dots, S\}$  be the set that contains the indices for each of the  $S$  species in the community. Suppose we take a sample (of individuals or plots) of size,  $\mu$ , from a collection of samples. Let the random variable  $\sigma_\mu$  be the subset of  $\sigma$  containing the indices of all species with non-zero abundance in the random sample. We define  $\langle \Phi_\mu \rangle$  to be,

$$\langle \Phi_\mu \rangle \equiv E \left[ \sum_{i>j} I_{\sigma_\mu}(i, j) d_{ij} \right]$$

where  $d_{ij}$  is a constant for all  $i$  and  $j$  on  $\sigma$  and  $I_{\sigma_\mu}(i, j)$  is a Bernoulli random variable (the indicator function) that equals 1 when both species  $i$  and  $j$  are in  $\sigma_\mu$  (and equals 0 otherwise). By the linearity of the expectation operator and the fact that  $I_{\sigma_\mu}(i, j)$  is a Bernoulli random variable,

$$\langle \Phi_\mu \rangle = \sum_{i>j} \text{Pr}(i, j \in \sigma_\mu) d_{ij}$$

where  $\text{Pr}(i, j \in \sigma_\mu)$  is the probability that species  $i$  and  $j$  are both in the sample,  $\sigma_\mu$ . This probability can be expanded in a manner that facilitates the search for an expression of  $\text{Pr}(i, j \in \sigma_\mu)$  in terms of the parameters of one of the four sampling models. Note that we can write  $\text{Pr}(i, j \in \sigma_\mu) = \text{Pr}(A \cap B)$  where  $A$  and  $B$  are the events that  $i$  and  $j$  respectively are on  $\sigma_\mu$  and  $\cap$  is the intersection symbol of elementary set theory. We need some further notation from elementary set theory:  $\cup$  is the union and  $A^c$  is the

complement of  $A$ . Using de Morgan’s laws (i.e. that  $(A \cup B)^c = A^c \cap B^c$  and  $(A \cap B)^c = A^c \cup B^c$ ), the inclusion-exclusion principle (i.e. that  $\text{Pr}(A \cup B) = \text{Pr}(A) + \text{Pr}(B) - \text{Pr}(A \cap B)$ ) and the fact that  $\text{Pr}(A) = 1 - \text{Pr}(A^c)$ , it is straightforward to show that,

$$\text{Pr}(A \cap B) = 1 - \text{Pr}(A^c) - \text{Pr}(B^c) + \text{Pr}(A^c \cap B^c).$$

Converting back into our previous notation, we have,

$$\begin{aligned} \text{Pr}(i, j \in \sigma_\mu) &= 1 - \text{Pr}(i \notin \sigma_\mu) - \text{Pr}(j \notin \sigma_\mu) \\ &\quad + \text{Pr}(i \notin \sigma_\mu, j \notin \sigma_\mu), \end{aligned}$$

where  $\text{Pr}(i \notin \sigma_\mu, j \notin \sigma_\mu)$  is the probability that neither  $i$  nor  $j$  are in the sample,  $\sigma_\mu$ . We can now write a general expression for  $\langle \Phi_\mu \rangle$ ,

$$\begin{aligned} \langle \Phi_\mu \rangle &= \Phi - \sum_{i>j} d_{ij} \\ &\quad \times [\text{Pr}(i \notin \sigma_\mu) + \text{Pr}(j \notin \sigma_\mu) - \text{Pr}(i \notin \sigma_\mu, j \notin \sigma_\mu)], \end{aligned}$$

where  $\langle \Phi_\mu \rangle \rightarrow \Phi$  as  $\mu \rightarrow \infty$  or when  $\mu$  reaches its maximum value. To derive the specific expressions for each type of FAD rarefaction (Table 1), we simply have to substitute the expressions for  $\text{Pr}(i \notin \sigma_\mu)$  and  $\text{Pr}(i \notin \sigma_\mu, j \notin \sigma_\mu)$  under each sampling model and for the limiting value,  $\Phi$ , of  $\langle \Phi_\mu \rangle$ .

## Appendix C. Bootstrap procedures

We now describe the bootstrap procedures that we used to approximate confidence intervals for rarefied functional diversity estimates in our Hubbard Brook bird community example. The procedure is similar for each diversity index under each sampling model and so we describe it for the general diversity index,  $D$ , and general sampling model. Bootstrapping was conducted by re-sampling  $v$  units with replacement from the  $v$  units that were sampled from the community and calculating the rarefied sample diversity,  $\langle \hat{D}_\mu \rangle_{\text{boot}}$ , of the bootstrap sample. This procedure was conducted  $B$  times and the standard deviation,  $\sigma_B$ , of the  $B$  rarefied sample diversities was calculated. The 95% confidence interval of the rarefied sample diversity,  $\langle \hat{D}_\mu \rangle$ , was calculated as  $\langle \hat{D}_\mu \rangle \pm 1.96\sigma_B$ . Note that this bootstrap standard deviation,  $\sigma_B$ , is not equivalent to taking the standard deviation of the iterations used in a Monte Carlo estimate of the rarefied sample diversity (Table 3). The former procedure yields an estimate of the standard error of the  $\langle \hat{D}_\mu \rangle$  while the latter does not. See Colwell et al. (2004) for a discussion of these issues in the context of species richness rarefaction and closed-form variance estimators.

For this bootstrap confidence interval to be valid, the distribution of the bootstrap samples,  $\langle \hat{D}_\mu \rangle_{\text{boot}}$ , must be approximately normal. This is true when the rarefied sample size,  $\mu$ , is small enough compared to the sample size,  $v$ . In our experience,  $\mu < (2/3)v$  is sufficiently small. However, the skew of the bootstrap distribution should be checked directly. Researchers can use the 2.5% and 97.5% values obtained from the bootstrap distribution to construct approximate confidence intervals if preferred. However, regardless of how the bootstrap confidence intervals are constructed, the rarefied sample size,  $\mu$ , should be chosen so that the bootstrap distribution is approximately symmetric; this is an important point. Recall that the bootstrap uses the observed distribution of data as an estimate of the true

population-level distribution. By sampling with replacement from the observed distribution, and calculating statistics of interest (such as diversity indices), we approximate the true sampling distribution of these statistics. This approximate sampling distribution can be used to compute approximate measures of error of these statistics such as standard errors and confidence intervals. These approximations get better as the observed distribution becomes more similar to the true population-level distribution of the data. In our situation, the diversity of a bootstrap sample can never exceed that of the diversity,  $\hat{D}$ , in the original sample. Therefore, when the rarefied sample size,  $\mu$ , equals the size,  $v$ , of the original sample, the upper bound of the bootstrap distribution will be equal to the statistic,  $\langle \hat{D}_\mu \rangle$ , being bootstrapped thus resulting in a skewed distribution (recall that  $\langle \hat{D}_\mu \rangle = \hat{D}$  when  $\mu = v$ ). This skew implies that the bootstrap approximation is failing because we would expect to obtain sample diversities higher than the diversity,  $\hat{D}$ , of our actual sample, upon repeated sampling from the population. The approximation improves, and the skew

disappears, as  $\mu$  becomes smaller because the expected diversity,  $\langle D_\mu \rangle$ , in a sample of size  $\mu$  is not likely to be higher than  $\hat{D}$  if  $\mu$  is small enough; in this case, it is reasonable to take  $\hat{D}$  as an upper-bound on the bootstrap distribution of  $\langle D_\mu \rangle$ . Therefore, the bootstrapping of rarefied diversities should only be used when the rarefied sample size,  $\mu$ , is sufficiently smaller than  $v$ .

We now describe a simple bootstrap method for approximating the confidence interval of the slope of a least-squares regression of a rarefied diversity index on a covariate of interest. Again, consider the generic diversity index,  $D$ , and sampling model. Draw  $B$  bootstrap re-samples,  $\langle \hat{D}_\mu \rangle_{\text{boot}}$ , for each community. Calculate the least-squares estimate of the slope,  $\hat{b}_{\text{boot}}$ , of the linear regression equation for each of the  $B$  re-samples. Let the standard deviation of the  $B$  re-sampled slopes be  $\sigma_B$ . Then the 95% confidence interval of the estimated slope,  $\hat{b}$ , is  $\hat{b} \pm 1.96\sigma_B$  or can be assessed directly from the 2.5% and 97.5% values obtained from the bootstrap if preferred.