1   # Understandable Multifunctionality Measures Using Hill Numbers

3   ## Abstract

4   In ecology, multifunctionality metrics measure the simultaneous performance of multiple
5   ecosystem functions. If species diversity describes the variety of species that together build the
6   ecosystem, multifunctionality attempts to describe the variety of functions these species perform.
7   A range of methods have been proposed to quantify multifunctionality, successively attempting
8   to alleviate problems that have been identified with the previous methods. This has led to a
9   proliferation of more-or-less closely related metrics which, however, lack an overarching
10   theoretical framework. Here we borrow from the comprehensive framework of species diversity
11   to derive a new metric of multifunctionality. Analogously to the effective number of species used
12   to quantify species diversity, the metric we propose is influenced both by the number of
13   functions as well as, crucially, the evenness of performance levels across functions. In addition,
14   the effective multifunctionality also considers the average level at which the functions are
15   performed. The result is a measure of the cumulative performance of the system were all
16   functions provided equally. The framework allows for the inclusion of the correlation structure
17   among functions, thus allowing it to account for non-independence between functions. We show
18   that the average metric is a special case of the newly proposed metric when all functions are
19   uncorrelated and performed at equal levels. We hope that by providing a new metric of
20   multifunctionality anchored in the rigorous framework of species diversity based on effective
21   numbers, we will overcome the considerable skepticism that the larger community of ecologists
22   has built against indices of multifunctionality. We thereby hope to help popularize this important
23   concept which, as biological diversity, describes a fundamental property of ecosystems and thus
24   lies at the heart of ecology.

Introduction

The past decade has witnessed the growth of the concept of ecosystem *multifunctionality*, defined typically as as a measure of the simultaneous performance of multiple functions. Having existed as a concept in the ecosystem service and land management literature for a long time (Hölting et al 2019) it has arisen, largely independently, in the biodiversity ecosystem function literature. From there, the concept has spread to community ecology writ large (Angelini et al. 2015), invasion biology (Ramus et al. 2017), land management (Nelson et al. 2009, Brandt et al. 2014, Binder et al. 2018), and more. The concept of multifunctionality is broad and can even be applied outside of community and ecosystem ecology - or even outside of ecology altogether. Its adoption as a unifying concept, however, has been slow - and for good reason.

Despite the popularity that the concept gained since it was first explicitly proposed in community ecology over a decade ago (Hector and Bagchi 2007, Gamfeldt et al. 2008), there is still no consensus on how multifunctionality should be quantified. A range of methods have been proposed where each subsequent method attempted to alleviate problems the authors had identified with the previous method, leading to a proliferation of metrics which cannot easily be compared. This state of affairs can lead to inconsistent interpretations of the results (Byrnes et al. 2014a, Gamfeldt and Roger 2017). We do note that this is a difficult problem. Throwing many different ecologists against this problem in a stimulating working group over multiple years (we are thankful for NCEAS for providing this venue) has likely resulted in more ink impregnated on foreheads hitting whiteboards than should be typical, to say nothing of others who have attempted to cut this Gordian knot before and since (Hector and Bagchi 2007, Brandt et al. 2014, Dooley et al. 2015, Rodríguez-Loinaz et al. 2015, Stürck and Verburg 2017, Manning et al.

48   2018, Meyer et al. 2018, Hölting et al. 2019). In part, the problem is due to a lack of an

49   underpinning body of theory describing how and where higher levels of multifunctionality

50   should arise. Regardless, this plurality of measurements has hampered a general understanding of

51   multifunctionality and possibly its adoption outside of the subfield of Biodiversity and

52   Ecosystem Function.

53

54   Species diversity, for a long time, suffered the same problems that multifunctionality suffers

55   today: there are an uncountable number of metrics to estimate alpha diversity, reaching from the

56   simple count of observed species without regard to their abundance (Richness), wide array of

57   metrics incorporating the relative abundance (Shannon entropy, Simpson index or the Berger-

58   Parker Dominance index to name some of the most common ones), to metrics based on the

59   histograms of the abundance distribution (e.g. Fisher's alpha) (Magurran and McGill 2010). Yet,

60   starting with MacArthur (1965), then Hill (1973), followed by Jost (2006), and most recently

61   Chao et al. (2014a, 2019), a common framework for species diversity has been developed based

62   on information theory. This framework, the effective number of species, encompasses the vast

63   majority of previous metrics and is able to handle a wide variety of different issues in diversity

64   estimation. We propose it can do the same for multifunctionality.

The current state of multifunctional affairs

66   The definition of multifunctionality, the simultaneous performance of multiple functions, (*sensu*

67   Byrnes et al. 2016) presents a challenge in creating a metric. How do we define a metric that

68   captures both the level of performance of a broad suite of functions as well as the distribution of

69   differences in performance among functions? Researchers have sought to capture this question in

70   four ways, after standardizing functions to similar levels in order to prevent apples-to-oranges

71   comparisons.

72   First, many have opted for simplicity and taking the average of all functions (e.g., Maestre et al.

73   2012). This approach is, at first glance, appealing, particularly as it provides a metric that can be

74   put on a y-axis while a predictor is on the x-axis. However, it sacrifices crucial information about

75   the system. An arithmetic mean tells us what level of functioning we would expect were we to

76   sample any one function at random. Consider two plots - one where all functions are similar and

77   performing at half their value and one where half of the functions are at their maximum while

78   half are absent. The averaging approach says that they are identical. Geometric averaging (e.g.,

79   Hensel and Silliman 2013) appears to get around this problem to some degree, as the geometric

80   mean is less than the arithmetic mean by a function of the variance of the observations (the two

81   means are the same when all observations are the identical). However, given its formulation, one

82   critically low function can dominate the measurement - e.g., if even one function is 0, the

83   geometric mean will be 0.

84   Second, we have metrics of the Multivariate Diversity Interactions framework (Dooley et al.

85   2015). This elegant framework allows us to tease apart the importance of correlations between

86   functions and the contribution of different drivers to simultaneous change in those functions. It

87   does not, however, provide a holistic metric of multifunctionality *per se*, much like the overlap

88   approach before it (Hector and Bagchi 2007). This approach from Hector and Bagchi is a

89   generalization of Sørenson-Dice overlap (Dice 1945, Sørensen 1948). It provides key

90   information on redundancy versus unique contributions of species. While it gives us rich

91   information about a system, it lacks holistic interpretability.

92   Last, we have the threshold approach (Gamfeldt et al. 2008). This approach defines the

93   "presence" of a function if it meets a given threshold of performance, and then sums up the

94   number of functions performing at or above that threshold. There are clear issues with the

95   arbitrariness of thresholds in the absence of relevant theory or management goals. To remedy

96   this, the multiple threshold approach (Byrnes et al. 2014a) seeks to balance the goals of

97   measuring the simultaneous performance of multiple ecosystem functions with the arbitrariness

98   of choosing a threshold of relevance for those functions. The results, however, can be difficult to

99   interpret. Looking at multiple relationships between diversity and threshold-based

100   multifunctionality does not provide a metric *per se*. While there are key metrics that can be

101   extracted from looking at change in slope over different thresholds, the meaning of these

102   quantities in the context of multifunctionality are not obvious. The values can vary dramatically

103   depending on the exact threshold chosen and on the method used to standardize the functions –

104   even for simulated communities where all functions perform identically (Gamfeldt and Roger

105   2017, Figure S1). In the absence of suitable null-models, and even if the approach yields rich

106   information about multifunctionality *sensu stricto*, it is unwieldy for most if not all who choose

107   to use it, and we have noticed that many of those who do, often report a single threshold in their

108   main text and leave further exploration to supplementary materials anyway. Related, more

109   recent, efforts have sought to use dimensionality-reducing techniques which have yielded metrics

110   that, while useful, similarly lack easily interpretable meaning (Meyer et al. 2018).

111   **Hill numbers and effective diversity**

112   Over the last 70 years, ecologists studying how to measure species diversity have shown that the

113   vast majority of previous diversity metrics can be organized into a common framework

114   (Macarthur 1965, Hill 1973, Jost 2006, Chao et al. 2014a, 2019). This is true for indices such as

115   Shannon entropy, all Simpson measures, all Renyi entropies, all HCDT or "Tsallis" entropies and

116   species richness (Jost, 2006) and more. All can be expressed as generalized entropies that can be

117   converted to an effective number of species of "order" q which specifies the weighting of

118   proportional abundances. The general formula for the diversity of order q for S species is the

119   following:

120
$$^qD = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)} \quad (1)$$

121   Here, p_i is the relative abundance of the ith species and q is the weight given to the species'

122   relative abundances. Species richness, the effective number of species based on Shannon

123   entropy, the effective number of species based on the Simpson index, and the Berger-Parker

124   dominance index are all effective numbers of species of order $q = 0$, 1, 2 and ∞, respectively.

125   (Note that the formula is undefined for $q = 1$, but its limit $q \to 1$ is $\exp(-\Sigma p_i \log p_i)$). The

126   effective number of species of order q is also often referred to as Hill numbers.

127

128   We propose leveraging this framework for a more meaningful and less *ad hoc* metric of

129   multifunctionality composed of two parts: (1) the effective *number* of functions that are

130   performed and (2) the *arithmetic mean performance* of the functions that are measured. Our

131   proposed multifunctionality index is then the product of both terms. This approach draws on

132   ideas already swirling in the multifunctionality literature (Brandt et al. 2014, Rodríguez-Loinaz

133   et al. 2015, Stürck and Verburg 2017, Hölting et al. 2019). With this approach, we aim to

134   provide a unifying framework for the measurement of multifunctionality.

Multifunctionality as the product of effective number of functions and average level of functions

To define the effective number of functions, we begin with a set of measurements on k functions (Table 1) that have been standardized to a common scale (i.e., between 0 and 1 where 0 means no function and 1 means maximum level of function). Let F_i , $i \in 1, 2, \ldots K$ show the level of function for function i (Table 1). The relative proportion a function contributes to the whole is defined as

$$p_i = \frac{F_i}{\sum F_i} \quad (2)$$

We can now substitute the relative proportion into the formula for the effective number of types given in Eq 1

$$^qN = \left(\sum_{i=1}^{K} p_i^q \right)^{1/(1-q)} \quad (3)$$

where ^qN is the **effective number of functions** for some order q (Table 1). *The effective number of functions here translates to* the equivalent number of functions were all functions provided at the same level. Effective number of functions tells us nothing about total level of functioning. Average function can be low or high (see below and Fig. 1). Rather, ^qN tells us how many functions we would see in an equivalent system where all functions were performing at the same level. This calculation requires a decision on how sensitive the function should be to differences in the level at which functions are provided, which is done by setting q (for lower values of q, differences in performance are less influential than for high values of q).

156   Adapting from Jost (2006), if $q = 0$, this is the total number of functions K, which is unimportant

157   as it is just the number of functions measured. In essence, all functions have equal weight

158   regardless of performance. For $q = 1$, the approximation of this function is equivalent to results

159   from Shannon diversity for species shown earlier – which we note has been used as a

160   multifunctionality metric previously (Stürck and Verburg 2017). For $q > 1$, functions performing

161   at higher levels are given greater weight. At $q = 2$, we get results that are equivalent to the

162   number of functions calculated from Simpson's diversity (inverse Simpson index). If one of our

163   goals is to up-weight high performing functions, $q = 2$ is a reasonable choice, while $q = 1$ is

164   sufficient as it accommodates information about unequal levels of functioning proportional to the

165   relative functional performance. Lower values of q would upweight low-performing functions -

166   which might be desirable in certain contexts when low levels of functions are sufficient. The

167   ability to modulate the sensitivity of the metric to high or low performing functions thus provides

168   a strong tool for both ecologists and managers. In the absence of a justification for a particular

169   value of q, exploring the robustness of results to different choices of q could prove fruitful as it

170   has in biodiversity research. Further exploration of how q relates to management goals or

171   ecological theory of multifunctionality would be a fruitful avenue for future research.

172

173   Effective number of functions does not accommodate knowledge about the absolute level of

174   functioning in a system - a true metric of multifunctionality. As long as the distribution of p_i is

175   the same, a system with high average function and low average function will look the same.

176   Indeed, for some values of q (e.g., $q = 1$), under some scenarios if one function goes up, the

177   effective number of functions can actually drop. To achieve the translation to a metric of

178   multifunctionality, we need to take into account the level at which the functions are performed:

179   the arithmetic mean of the function values standardized to a common scale, which we define as A

180   (Table 1). As we are using standardized values as before, A will range from 0 to 1.

181

182   We can then calculate **effective multifunctionality** of order q (Table 1) as the product of both

183   terms. We remind readers that A is an expected value – it provides information on the expected

184   level of one function sampled at random from the cluster of functions. Scaling A by ^{q}N gives a

185   metric of multifunction summed across the suite of functions – *the cumulative performance of*

186   *the system were it composed of functions all performing at equal levels*

187
$$^{q}M_{ef} \ = \ ^{q}NA \quad (4)$$

188   This metric, where $^{q}M_{ef}$ is **effective multifunctionality** for order q, will have a maximum value

189   of K, the total number of functions measured in the system, as maximum performance is all

190   functions performing at a standardized level of 1. Alternatively, we can standardize by the total

191   number of functions to calculate the fraction of performance achieved by the whole system, of

192   **standardized effective multifunctionality**, ^{q}M (Table 1), which could facilitate future

193   comparisons among studies.

194
$$^{q}M \ = \ ^{q}NA/K \quad (5)$$

195   Given that we have scaled by K, this metric measures the level of individual functions in an

196   equivalent system where all functions have the same level of performance. However, care has to

197   be taken when comparing multifunctionality values across systems that measure different sets of

198   functions and the value of such comparisons is a topic that eludes consensus even among authors

199   of this article.

200

201   Why must we consider the level of function and effective number of functions together in one

202   metric? First, when A is less than one, multiple values of ^qN are possible depending on the

203   distribution of performance of functions. Second, the average functioning and the effective

204   number of functions are not independent of each other. The upper limit of ^qN is K by definition.

205   ^qN is K when all functions are performing at the same level, i.e. $E[A] = F_1, F_2, F_3...F_K$. This can

206   be achieved for any value of A, low or high. Thus, ^qN has no information about the level of

207   function achieved. The lower limit of ^qN (as long as q is not 0) occurs at the maximum level of

208   unevenness of F across all functions. This lower limit will vary by different levels of A. For

209   example, ^qN will always be K when A is 1. Similarly, if all functions but one have a value of 0,

210   then ^qN will be 1, as p_i for the non-zero function will be 1. If more than 1 function exceeds 0,

211   then the lower limit is set by a combination of A and the number of dominant functions, i.e., the

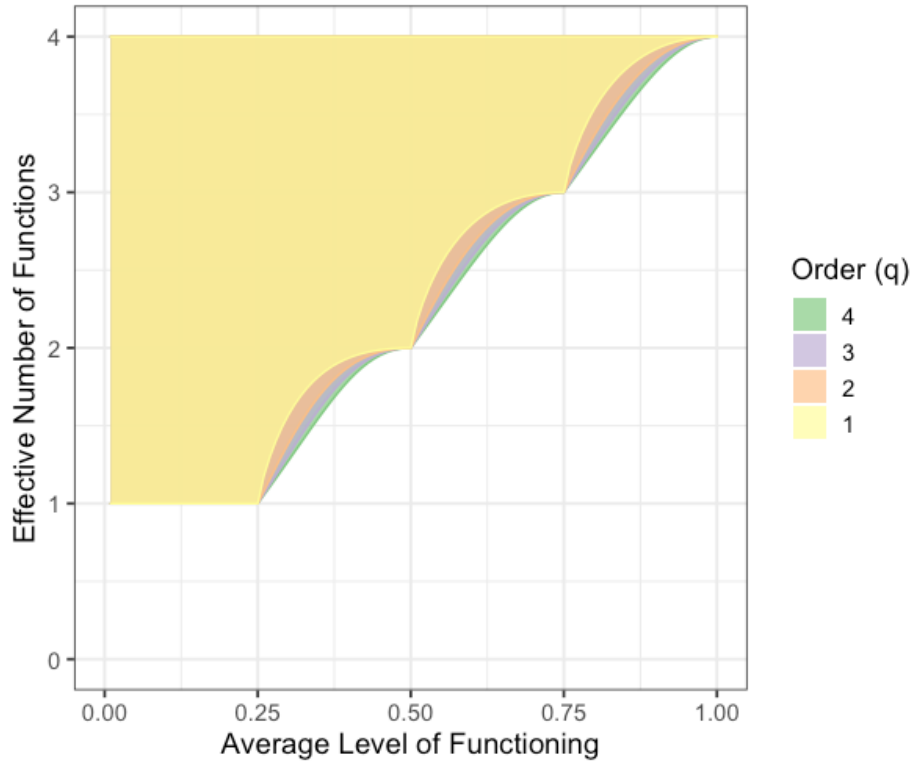212   functions performing at $F_i = 1$ (Figure 1). See Appendix 1 for derivation.

213

Figure 1. Relationship between effective number of functions, ^qN and average level of functions, A, for 4

functions at $q = 1,2,3,$ and 4. Filled areas show the full range of possible values of effective number of

functions, ^qN, under different values of average level of function, A. Note, areas with a higher order (q) do

overlap completely with lower orders, save for the additional edge that is visible.

Most importantly, a combined metric satisfies our definition of multifunctionality. High numbers

imply both a high level of function and high functional evenness (i.e., p_i is close to $1/K$ for all

functions). Low values imply that, even if a single function is being optimized, the assemblage of

functions as a whole is not performing at a high level. The relationship can also be easily

decomposed into its constituent parts for a more detailed examination of its behavior.

226 Last, from a convenience standpoint, having a single metric allows us to begin to examine it as

227 any other response variable. In the Biodiversity and Ecosystem Functioning world, we might

228 look at additive partitioning in addition to complementary overlap approaches. In global change

229 biology, we can look at the stability, resistance, and resilience of this metric in ecosystems

230 confronting human stressors. This metric can be used just as any other univariate metric in any

231 field, leading to easy adoption of the multifunctionality concept across many fields of endeavor.

232 The options are open.

233

A note on standardization

235 As pointed out in discussions of multifunctionality, how functions are standardized matters

236 (Gamfeldt and Roger 2017, Manning et al. 2018). First is the choice of direction - what implies

237 positive function? When this approach is applied to ecosystem services, the choice of direction is

238 hopefully a straightforward choice - although consider tradeoffs between nutrient cycling rates

239 and storage as a tricky context. Further, not all functions are equally important, particularly in the

240 case of contributions to services. Fortunately, choosing an 'optimal' level of function to link to 1

241 can alleviate this (e.g., if 25% of function is sufficiently high for the provision of a service, 25%

242 or higher can be considered a '1'). Functions can also be upweighted or downweighted in the

243 calculation of ^{q}N so long as $\sum p_i = 1$. Choices for standardization are often best made in the

244 context of a specific system or application and must be transparently justified.

245

Correlated functions and multifunctionality

247   A great deal of debate in the multifunctionality literature has sprung from the issue of how to

248   deal with correlations between functions (Bradford et al. 2014a, b, Byrnes et al. 2014b, Manning

249   et al. 2018). As with species diversity, the rationale is that not all functions are equally different

250   and that a metric of true multifunctionality should identify "variables that represent independent

251   aspects of ecosystem functioning" (Manning et al. 2018). As an illustration, consider a scenario

252   where we measure a set of functions, several of which result from a shared mechanism so that

253   they are inextricably linked (for example, growth rate and final biomass). If we want to study the

254   circumstances under which overall multifunctionality is maximized, the results will be

255   disproportionately driven by circumstances maximizing this mechanism influencing the set of

256   correlated functions; thus a naive application of a multifunctionality measure would implicitly

257   upweight the importance of this function cluster over uncorrelated functions.

258

259   In the Ecosystem service literature, this has led to the concept of Ecosystem service bundles

260   (Raudsepp-Hearne et al. 2010), where cluster analysis is used to identify groups (or bundles) of

261   ecosystem services that tend to occur together. Following the same logic, Manning et al. (2018)

262   proposed performing a cluster analysis of ecosystem functions and building a dendrogram

263   representing the distance matrix between functions. The performance of single functions is then

264   averaged within clusters and multifunctionality is calculated among clusters and not individual

265   functions. This is conceptually very similar to metrics of phylogenetic and functional species

266   diversity that construct phylogenetic trees or dendrograms based on functional similarities and

267   measure species diversity taking into account the dendrogram structure (e.g. Faith's PD, Allen's

268   H, Petchey and Gaston's FD, Rao's Quadratic entropy etc.). The problem is that, as discussed by

269   Chao et al. (2014a), dendrograms are very sensitive to the type of clustering method used and the

270   shape of a dendrogram can vary substantially between different methods (Poos et al. 2009). For

271   the method proposed by Manning et al 2018, it is also unclear at what depth the dendrogram

272   should be cut (i.e., what clusters are considered as independent). Fortunately, extensions of the

273   Hill numbers framework to account for species similarity, for example in terms of traits or

274   genetic relatedness (Chao et al. 2014a, 2019), provide a solution to the challenge of evaluating

275   multifunctionality across correlated functions.

276

277   To account for correlation among functions in the effective multifunctionality framework, we

278   need to incorporate several concepts from Chao et al. (2019). First, how similar are functions?

279   For this, we need a distance matrix of some sort (Table 1). If this is based on *a priori*

280   mechanistic information about the underlying processes driving functions (e.g., physiological or

281   evolutionary constraints), so much the better. We often do not have such a matrix. Other options

282   include a matrix derived from principal components or other methods of constructing distance

283   matrices (see Manning et al. 2018). In the absence of such information, a practical choice could

284   be to look at the correlation matrix among functions, R. Defining $D = (1 - R)/2$ would create a

285   distance matrix where a d_{ij} value of 0 means two functions are inextricably linked (perfectly

286   correlated) while 1 means they trade-off completely ($r_{ij} = -1$). This is one choice. We note that

287   some functions could be correlated for non-biological reasons, and as such a biologically-based

288   distance matrix might be wiser. We suggest that determining the proper way to create a distance

289   matrix between functions for the estimation of effective diversity is an exciting area of research.

290

291   With a distance matrix in hand, we need to ask ourselves, how distinct must two functions be

292   before we consider them completely different? This incorporates the threshold τ (Table 1)

293   proposed in Chao et al. (2019). Using this threshold, the distance matrix is truncated such that

294    $d_{ij}(\tau) = \min(d_{ij}, \tau)$. If, for example, $\tau \leq \min(d_{ij})$, all functions are equally distinct and we recover

295   the results ignoring correlations. If $\tau = 1$ or the maximum distance in the matrix, we fully

296   incorporate information from the distance matrix with no truncation. If $\tau = 0.5$, all negatively

297   correlated functions are as distinct as uncorrelated functions. Chao et al. (2019) recommend

298   using the average weighted pairwise distance between any two functions as τ

299
$$d_{mean} = \sum_{i=1}^{K} \sum_{j=1}^{K} d_{ij} p_i p_j \quad (6)$$

300   as setting τ to d_{mean} yields results that are consistent with comprehensive evaluations of all

301   possible values of τ.

302

303   We can then use our truncated distance matrix, $d_{ij}(\tau)$, to define the effective number of

304   underlying functions, $^q N_u$ as follows when we are interested in the effective number of

305   underlying "functions". This equation takes advantage of $d_{ij}(\tau)/\tau$ signifying the portion of one

306   function that is distinct from another.

307

308
$$^q N_u = \left(\sum_{i=1}^{K} p_i \left(\sum_{j=1}^{K} \left[1 - \frac{d_{ij}(\tau)}{\tau} \right] p_j \right)^{q-1} \right)^{1/(1-q)} \quad (7)$$

309

310   In essence, the inner part of the equation is the proportion of function i multiplied by the summed

311   portion of all functions that contribute to the same underlying function. For $q = 1$, we need to use

312   the limit, as before, so

313
$$^qN_u = exp\left(-\sum_{i=1}^{K} p_i \ log\left(\sum_{j=1}^{K}\left[1 - \frac{d_{ij}(\tau)}{\tau}\right]p_j\right)\right) \quad (8)$$

314   This effective number of functions can then be used to calculate multifunctionality as above. The

315   average, A, does not need to be modified, because correlations do not systematically bias

316   estimates of the mean (although uncertainty in that mean could be biased downwards by

317   correlations). The correlations only need to be accounted for in the estimate of ^qN.

318

319   **Discussion**

320   With the framework above in hand, we are hopeful for a proliferation of literature investigating

321   both new and old ideas in multifunctionality. With its ties to theoretically grounded methods of

322   measuring species diversity and its flexibility to encompass some of the real sticking points in

323   the field, we hope that Ecology will embrace this framework moving forward. It produces more

324   intuitive metrics - effective number of functions performing at equal levels, effective system-

325   wide multifunction – number of functions scaled by total level of performance, and standardized

326   versions of both - as well as offering a solution to account for correlated functions. To aid in this

327   advance, see Appendix 2 where we present a worked example in R (R Core Development Team

328   2008). Over the last decade and a half of its development, the multifunctionality literature has

329   grown more slowly than it should, in no small part due to the *ad hoc* nature of the metrics we

330   have developed. We hope that period is at an end.

331

332   Further, the metrics presented here are the foundation of a much larger framework that has seen

333   deep exploration in the species diversity world. By embracing this framework for

334   Multifunctionality research, we open up new vistas for ecology. Some are small - what the

335   consequences and best choices for τ are, how we calculate distance matrices, at what scale to

336   evaluate correlation between functions, should we incorporate changes in correlation and

337   distance matrices as an ecological response in and of themselves, how we standardize functional

338   measurements, and many more. Other new areas of inquiry are quite large. Can we use this

339   framework to begin to address the problem of unmeasured functions as we deal with unmeasured

340   species (Chao et al. 2014b)? Can we use it to think about turnover in multifunctionality across

341   space and time as we do with beta diversity (Chao and Ricotta 2019) or think about partitioning

342   landscape multifunctionality into different components (Jost 2007)? Particularly as we think

343   about global change's impacts on ecosystem services, and not just functions, at multiple scales,

344   this framework lays out fruitful avenues for future exploration.

345

346   What does our approach mean for previous metrics? In short, existing approaches have addressed

347   specific aspects of measuring multifunctionality, but are limited in their scope. Here we provide

348   a metric that addresses the generalized measurement of multifunctionality that meets its

349   definition – the "simultaneous performance of multiple functions" (Byrnes et al. 2016) – in a

350   way not fully encompassed by previous metrics. The averaging metric comes closest to what is

351   presented here, as it is a special case where $q = 0$. In the supplementary materials we provide an

352   example using data from Duffy et al. 2003 and the *multifunc* package (v 0.9.4) in R, now

353   including metrics in this paper, that shows some concordance between the two. The threshold

354   approach seeks to remedy some of the drawbacks of using $q = 0$, and can be useful for managers

355   that have specific targets of functional performance they need to meet. However, it is too

356   sensitive to a number of decisions, as well as complicated to interpret (Gamfeldt and Roger

357   2017). Other metrics in the literature are often targeted at questions related to multifunctionality,

358   but do not directly address a univariate measure of simultaneous function (e.g., the approach of

359   Dooley et al. 2015 seeks to look at a multivariate response and quantify tradeoffs rather than

360   provide a single univariate measure). Rather, we have sought to provide a metric that builds on

361   past work while providing a robust foundation for the future of multifunctionality research.

362

363   Ultimately, we feel that the proliferation of univariate multifunctionality metrics without strong

364   theoretical underpinning has caused a great deal of confusion about how to measure

365   multifunctionality. We hope that this piece will provide the field of multifunctionality with a way

366   out of its current state of division and confusion. Further, we hope it provides food for additional

367   theory that addresses the causes and consequences of ecosystem multifunctionality, something

368   that is currently sorely lacking but highly relevant to policy and management (e.g., the efforts of

369   the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services). We

370   have been heartened by the idea leaving the cradle of the field of biodiversity and ecosystem

371   function, and feel that it has the promise to provide a holistic unifying concept for anyone

372   interested in capturing a snapshot of system dynamics in a single meaningful metric with direct

373   ties to the beautifully developing field of diversity partitioning. Much is to be done on honing the

374   particulars of this approach, but we feel it offers a strong theory-driven unified approach that will

375   enable the field of Multifunctionality research to move forward swiftly.

376

379

380   **Data Access**

381   --Redacted for review, but see supplementary materials for all relevant code--

382

383   **References**

384   Angelini, C., van der Heide, T., Griffin, J. N., Morton, J. P., Derksen-Hooijberg, M., Lamers, L.

385   P. M., Smolders, A. J. P. and Silliman, B. R. 2015. Foundation species' overlap enhances

386   biodiversity and multifunctionality from the patch to landscape scale in southeastern

387   United States salt marshes. - Proc. R. Soc. B Biol. Sci. 282: 20150421.

388   Binder, S., Isbell, F., Polasky, S., Catford, J. A. and Tilman, D. 2018. Grassland biodiversity can

389   pay. - PNAS: 201712874.

390   Bradford, M. A., Wood, S. A., Bardgett, R. D., Black, H. I. J., Bonkowski, M., Eggers, T.,

391   Grayston, S. J., Kandeler, E., Manning, P., Setälä, H. and Jones, T. H. 2014a.

392   Discontinuity in the responses of ecosystem processes and multifunctionality to altered

393   soil community composition. - PNAS 111: 14478–14483.

394   Bradford, M. A., Wood, S. A., Bardgett, R. D., Black, H. I. J., Bonkowski, M., Eggers, T.,

395   Grayston, S. J., Kandeler, E., Manning, P., Setälä, H. and Jones, T. H. 2014b. Reply to

396   Byrnes et al.: Aggregation can obscure understanding of ecosystem multifunctionality. -

397   PNAS 111: E5491–E5491.

398   Byrnes, J. E. K., Gamfeldt, L., Isbell, F., Lefcheck, J. S., Griffin, J. N., Hector, A., Cardinale, B.

399   J., Hooper, D. U., Dee, L. E. and Emmett Duffy, J. 2014a. Investigating the relationship

400   between biodiversity and ecosystem multifunctionality: challenges and solutions. -

401   Methods Ecol. Evol. 5: 111–124.

402   Byrnes, J., Lefcheck, J. S., Gamfeldt, L., Griffin, J. N., Isbell, F. and Hector, A. 2014b.

403   Multifunctionality does not imply that all functions are positively correlated. - PNAS

404   111: E5490.

405 Chao, A. and Ricotta, C. 2019. Quantifying evenness and linking it to diversity, beta diversity,

406   and similarity. - Ecology 100: e02852.

407 Chao, A., Chiu, C.-H. and Jost, L. 2014a. Unifying Species Diversity, Phylogenetic Diversity,

408   Functional Diversity, and Related Similarity and Differentiation Measures Through Hill

409   Numbers. - Annu. Rev. Ecol. Evol. Syst. 45: 297–324.

410 Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. and Ellison, A. M.

411   2014b. Rarefaction and extrapolation with Hill numbers: a framework for sampling and

412   estimation in species diversity studies. - Ecol. Monogr. 84: 45–67.

413 Chao, A., Chiu, C.-H., Villéger, S., Sun, I.-F., Thorn, S., Lin, Y.-C., Chiang, J.-M. and Sherwin,

414   W. B. 2019. An attribute-diversity approach to functional diversity, functional beta

415   diversity, and related (dis)similarity measures. - Ecol. Monogr. 89: e01343.

416 Dice, L. R. 1945. Measures of the Amount of Ecologic Association Between Species. - Ecology

417   26: 297–302.

418 Dooley, Á., Isbell, F., Kirwan, L., Connolly, J., Finn, J. A. and Brophy, C. 2015. Testing the

419   effects of diversity on ecosystem multifunctionality using a multivariate model (Y

420   Buckley, Ed.). - Ecol. Lett. 18: 1242–1251.

421 Gamfeldt, L. and Roger, F. 2017. Revisiting the biodiversity–ecosystem multifunctionality

422   relationship. - Nat. Ecol. Evol. 1: 1–7.

423 Gamfeldt, L., Hillebrand, H. and Jonsson, P. R. 2008. Multiple functions increase the importance

424   of biodiversity for overall ecosystem functioning. - Ecology 89: 1223–1231.

425 Hector, A. and Bagchi, R. 2007. Biodiversity and ecosystem multifunctionality. - Nature 448:

426   188–190.

427   Hensel, M. J. S. and Silliman, B. R. 2013. Consumer diversity across kingdoms supports

428   multiple functions in a coastal ecosystem. - PNAS 110: 20621–20626.

429   Hill, M. O. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. - Ecology

430   54: 427–432.

431   Hölting, L., Beckmann, M., Volk, M. and Cord, A. F. 2019. Multifunctionality assessments –

432   More than assessing multiple ecosystem functions and services? A quantitative literature

433   review. - Ecol. Indic. 103: 226–235.

434   Jost, L. 2006. Entropy and diversity. - Oikos 113: 363–375.

435   Jost, L. 2007. Partitioning Diversity into Independent Alpha and Beta Components. - Ecology

436   88: 2427–2439.

437   Macarthur, R. H. 1965. Patterns of Species Diversity. - Biol. Rev. 40: 510–533.

438   Maestre, F. T., Quero, J. L., Gotelli, N. J., Escudero, A., Ochoa, V., Delgado-Baquerizo, M.,

439   Garcia-Gomez, M., Bowker, M. A., Soliveres, S., Escolar, C., Garcia-Palacios, P.,

440   Berdugo, M., Valencia, E., Gozalo, B., Gallardo, A., Aguilera, L., Arredondo, T., Blones,

441   J., Boeken, B., Bran, D., Conceicao, A. A., Cabrera, O., Chaieb, M., Derak, M., Eldridge,

442   D. J., Espinosa, C. I., Florentino, A., Gaitan, J., Gatica, M. G., Ghiloufi, W., Gomez-

443   Gonzalez, S., Gutierrez, J. R., Hernandez, R. M., Huang, X., Huber-Sannwald, E., Jankju,

444   M., Miriti, M., Monerris, J., Mau, R. L., Morici, E., Naseri, K., Ospina, A., Polo, V.,

445   Prina, A., Pucheta, E., Ramirez-Collantes, D. A., Romao, R., Tighe, M., Torres-Diaz, C.,

446   Val, J., Veiga, J. P., Wang, D. and Zaady, E. 2012. Plant Species Richness and

447   Ecosystem Multifunctionality in Global Drylands. - Science 335: 214–218.

448   Magurran, A. E. and McGill, B. J. 2010. Biological Diversity: Frontiers in Measurement and

449   Assessment. - OUP Oxford.

450 Manning, P., van der Plas, F., Soliveres, S., Allan, E., Maestre, F. T., Mace, G., Whittingham, M.

451   J. and Fischer, M. 2018. Redefining ecosystem multifunctionality. - Nat. Ecol. Evol. 2:

452   427–436.

453 Meyer, S. T., Ptacnik, R., Hillebrand, H., Bessler, H., Buchmann, N., Ebeling, A., Eisenhauer,

454   N., Engels, C., Fischer, M., Halle, S., Klein, A.-M., Oelmann, Y., Roscher, C., Rottstock,

455   T., Scherber, C., Scheu, S., Schmid, B., Schulze, E.-D., Temperton, V. M., Tscharntke,

456   T., Voigt, W., Weigelt, A., Wilcke, W. and Weisser, W. W. 2018. Biodiversity–

457   multifunctionality relationships depend on identity and number of measured functions. -

458   Nat. Ecol. Evol. 2: 44–49.

459 Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, Dr., Chan, K. M., Daily,

460   G. C., Goldstein, J., Kareiva, P. M., Lonsdorf, E., Naidoo, R., Ricketts, T. H. and Shaw,

461   Mr. 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity

462   production, and tradeoffs at landscape scales. - Front. Ecol. Environ. 7: 4–11.

463 Poos, M. S., Walker, S. C. and Jackson, D. A. 2009. Functional-diversity indices can be driven

464   by methodological choices and species richness. - Ecology 90: 341–347.

465 R Core Development Team 2008. R: A Language and Environment for Statistical Computing.

466 Ramus, A. P., Silliman, B. R., Thomsen, M. S. and Long, Z. T. 2017. An invasive foundation

467   species enhances multifunctionality in a coastal ecosystem. - PNAS 114: 8580–8585.

468 Raudsepp-Hearne, C., Peterson, G. D. and Bennett, E. M. 2010. Ecosystem service bundles for

469   analyzing tradeoffs in diverse landscapes. - Proceedings of the National Academy of

470   Sciences 107: 5242–5247.

471 Rodríguez-Loinaz, G., Alday, J. G. and Onaindia, M. 2015. Multiple ecosystem services

472   landscape index: A tool for multifunctional landscapes conservation. - J. Environ.

473   Manage. 147: 152–163.

474 Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based

475   on similarity of species content and its application to analyses of the vegetation on

476   Danish commons. - Biol. Skr. 5: 1–34.

477 Stürck, J. and Verburg, P. H. 2017. Multifunctionality at what scale? A landscape

478   multifunctionality assessment for the European Union under conditions of land use

479   change. - Landsc. Ecol. 32: 481–500.

480

481 **Table 1:** Definitions of variables used to calculate and define the effective multifunctionality of
482 an ecosystem.

	Variable	Description
Variables that define elements of effective multifunctionality	$^qM_{ef}$	Effective multifunctionality of order q. The cumulative performance of the system were it composed of functions all performing at equal levels.
	^qM	Standardized effective multifunctionality of order q. Expresses effective multifunctioning as a fraction of the maximum attainable value when all functions are performing at their maximum (i.e., all functions are 1). It is obtained by dividing $^qM_{ef}$ by k to provide a metric comparable across systems, although care is required when comparing different sets of functions.
	^qN	Effective number of functions of order q. Describes the equivalent number of functions were all functions provided at the same level.
	^qN_u	Correlation-corrected effective number of functions of order q. Quantifies the effective number of functions were all functions independent and provided at the same level.
Variables used to calculate effective multifunctionality	k	The number of functions measured.
	F_i	Level of function i, standardized by dividing by its maximum observed or theoretical value.
	p_i	Proportion of summed functions provided by function i.
	q	Order of diversity (default = 1). Smaller values increase weight of lower performing functions, so that ^0D is the number of functions provided in any amount. Larger values express the degree to which high-performing functions are upweighted.
	A	The arithmetic mean of the k measured functions.

D		A dissimilarity matrix describing the independence of functions from each other. Based on *a priori* mechanistic information (e.g., physiological or evolutionary constraints) or empirical estimates of among functions correlation matrix, R ($D = 0.5(1 - R)$). Empirical estimates of R can be obtained from the data set being analyzed or additional data.
τ		Threshold dissimilarity between two functions for them to be considered independent in the calculation of ^qN_u. The dissimilarity matrix is truncated, so that any values above τ are converted to 1. The average weighted pairwise distance among functions, d_{mean} is the recommended default, following Chao et al. (2019). If $\tau = 1$, no truncation occurs; if $\tau = 0.5$, all negatively correlated functions are considered independent. If $\tau = 0$, all functions are treated as completely independent and we recover the value of ^qN uncorrected for correlations among functions.

483