

## Variation in the methods leads to variation in the interpretation of biodiversity-ecosystem multifunctionality relationships

Xin Jing<sup>1,2\*#</sup>, Case M. Prager<sup>1,3</sup>, Aimée T. Classen<sup>1,2</sup>, Fernando T. Maestre<sup>4,5</sup>, Jin-Sheng He<sup>6</sup>, Nathan J. Sanders<sup>2,3,7</sup>

<sup>1</sup>Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT 05405, USA

<sup>2</sup>Gund Institute for Environment, University of Vermont, Burlington, VT05405, USA

<sup>3</sup>Center for Macroecology, Evolution, and Climate, Natural History Museum of Denmark, Copenhagen, Denmark

<sup>4</sup>Departamento de Biología, Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Móstoles, 28933, Spain

<sup>5</sup>Departamento de Ecología and Instituto Multidisciplinar para el Estudio del Medio “Ramon Margalef”, Universidad de Alicante, Alicante, Spain

<sup>6</sup>Department of Ecology, College of Urban and Environmental Sciences, Peking University, Beijing, China

<sup>7</sup>Environmental Program, Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT 05405, USA

**\* Corresponding author:**

Xin Jing, Tel: +32-16321795; E-mail: [xin.jing@kuleuven.be](mailto:xin.jing@kuleuven.be)

<sup>#</sup>Current Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E, 3001 Leuven, Belgium

This article has been accepted For publication And undergone full peer review but has Not been through the copyediting, typesetting, pagination And proofreading process, which may lead to differences between this

Version And the Version of Record. Please cite this article as doi: 10.1093/jpe/rtaa031.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Institute of Botany, Chinese Academy of Sciences and the Botanical Society of China. All rights reserved. For permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## **Abstract**

### ***Aims***

Biodiversity is often positively related to the capacity of an ecosystem to provide multiple functions simultaneously (*i.e.*, multifunctionality). However, there is some controversy over whether biodiversity-multifunctionality relationships depend on the number of functions considered. Particularly, investigators have documented contrasting findings that the effects of biodiversity on ecosystem multifunctionality do not change or increase with the number of ecosystem functions. Here, we provide some clarity on this issue by examining the statistical underpinnings of different multifunctionality metrics.

### ***Methods***

We used simulations and data from a variety of empirical studies conducted across spatial scales (from local to global) and biomes (temperate and alpine grasslands, forests and drylands). We revisited three methods to quantify multifunctionality including the averaging approach, summing approach and threshold-based approach.

### ***Important findings***

Biodiversity-multifunctionality relationships either did not change or increased as more functions were considered. These results were best explained by the statistical underpinnings of the averaging and summing multifunctionality metrics. Specifically, by averaging the individual ecosystem functions, the biodiversity-multifunctionality relationships equal the population mean of biodiversity-single function relationships, and thus will not change with the number of functions. Likewise, by summing the individual ecosystem functions, the strength of biodiversity-multifunctionality relationships increases as the number of functions increased. We proposed a scaling standardization method by converting the averaging or

summing metrics into a scaling metric, which would make comparisons among different biodiversity studies. In addition, we showed that the range-relevant standardization can be applied to the threshold-based approach by solving for the mathematical artefact of the approach (*i.e.*, the effects of biodiversity may artificially increase with the number of functions considered). Our study highlights different approaches yield different results and that it is essential to develop an understanding of the statistical underpinnings of different approaches. The standardization methods provide a prospective way of comparing biodiversity-multifunctionality relationships across studies.

**Keywords:** Averaging approach, Biodiversity, Ecosystem multifunctionality, Multiple threshold approach, Plant species richness, Spatial scale

Accepted Manuscript

*“The challenge remains to develop multifunctionality indices that appropriately account for the aggregate effects of contrasting individual functions when their responses depend on multiple drivers that vary in their effects either in space or time.” – Bradford et al. (2014a)*

Accepted Manuscript

## INTRODUCTION

Over the past decade, a key question in ecology has been whether biodiversity increases the capacity of an ecosystem to provide multiple functions simultaneously (*i.e.*, multifunctionality) (Gamfeldt *et al.*, 2008, Hector *et al.*, 2007, Hines, 2019, Manning *et al.*, 2018). In general, as biodiversity increases, so does at least one ecosystem function – typically aboveground net primary productivity (ANPP), *e.g.*, Hector *et al.* (1999), Hooper *et al.* (2012), Tilman *et al.* (2014). However, an ecosystem performs a variety of functions in addition to ANPP, and indeed there is an emerging body of research that seeks to estimate multifunctionality and to understand how biodiversity influences multifunctionality (Byrnes *et al.*, 2014b, Gotelli *et al.*, 2011, Maestre *et al.*, 2012, Manning, *et al.*, 2018, Zavaleta *et al.*, 2010). While many biodiversity-multifunctionality studies provide evidence that increasing biodiversity leads to higher multifunctionality (Fanin *et al.*, 2018, Hautier *et al.*, 2018, Lefcheck *et al.*, 2015a), such biodiversity effects become increasingly complex in naturally assembled communities (Snelgrove *et al.*, 2014, van der Plas, 2019). This pattern likely emerges because the mechanisms underlying biodiversity-multifunctionality relationships are not well understood when the number of functions considered increases (Fanin, *et al.*, 2018, Gamfeldt *et al.*, 2017, van der Plas *et al.*, 2016). For example, trade-offs among ecosystem functions could accumulate with the number of functions (Butterfield *et al.*, 2016, Byrnes *et al.*, 2014a, Dooley, 2018, Lefcheck, *et al.*, 2015a, Meyer *et al.*, 2018), thereby offsetting the effects of biodiversity on multifunctionality (Byrnes, *et al.*, 2014b, Zavaleta, *et al.*, 2010). Alternatively, the rate of multifunctionality loss as biodiversity is lost may increase because the levels of redundancy across ecosystem functions become lower when more functions are considered (Delgado-Baquerizo *et al.*, 2016a, Gamfeldt, *et al.*, 2008, Miki *et al.*, 2014, Mori *et al.*, 2016).

Currently, there are at least four ways to estimate multifunctionality (Byrnes, *et al.*, 2014b, Dooley *et al.*, 2015, Hölting *et al.*, 2019, Manning, *et al.*, 2018): the single functions approach, averaging approach, turnover approach and threshold-based approach (single vs. multiple thresholds) (see summaries by Byrnes, *et al.*, 2014b, Dooley, *et al.*, 2015).

Investigators might use the single functions approach if they desire to evaluate whether high diversity communities maintain more individual functions with higher values than do less diverse communities (Duffy *et al.*, 2003). The averaging approach might be used if researchers want to compare the average effect of diversity on multiple ecosystem functions (Hooper *et al.*, 1998, Maestre, *et al.*, 2012). The turnover approach might be used if the goal is to assess whether different species contribute different ecosystem functions or services (Hautier, *et al.*, 2018, Hector and Bagchi, 2007, Isbell *et al.*, 2011, Wagg *et al.*, 2019). And finally, the threshold-based approach might be used if the aim is to compare whether a community can simultaneously achieve high levels of multifunctionality for a given threshold or multiple thresholds (Byrnes, *et al.*, 2014b, Gamfeldt, *et al.*, 2008, Zavaleta, *et al.*, 2010).

Although these four approaches provide useful information about biodiversity-multifunctionality relationships, multifunctionality is not just a quantitative metric. Indeed, it is also applied to quantify the overall performance of ecosystem quality, *i.e.*, the extent to which an ecosystem has desirable properties from the viewpoint of human society (Allan *et al.*, 2015, Manning, *et al.*, 2018, Slade *et al.*, 2017, Song *et al.*, 2020). The problem with multiple metrics/valuations of multifunctionality is that it hinders synthesis across studies or comparing effects of particular sub-components of ecosystems (Byrnes, *et al.*, 2014b).

Many studies have demonstrated that the positive contribution of species to multifunctionality increases with the number of functions (e.g., Hautier, *et al.*, 2018, Hector and Bagchi, 2007, Isbell, *et al.*, 2011, Meyer, *et al.*, 2018). Recently, three studies have explored whether biodiversity-multifunctionality relationships are contingent on the number

of functions considered. Using simulation models, Gamfeldt and Roger (2017) found that increasing the number of functions did not change biodiversity-multifunctionality relationships. That is, the effects of biodiversity on multifunctionality are approximately identical to the average effects of biodiversity on individual ecosystem functions. The results were supported by an empirical study examining biodiversity-multifunctionality relationships on Swedish islands (Fanin, *et al.*, 2018). However, a study conducted in a German grassland found different trends (Meyer, *et al.*, 2018). Specifically, as more functions were included in the estimate of multifunctionality, biodiversity-multifunctionality relationships became increasingly positive. These contrasting findings, obtained using different modelling and experimental approaches in different study ecosystems, highlight the need for additional and more detailed analyses of how the number of functions influences biodiversity-multifunctionality relationships.

Here we use simulation models and empirical data to show that the statistical underpinnings of quantitative multifunctionality metrics are important for understanding whether increasing the number of functions will alter biodiversity-multifunctionality relationships. We first re-examine and expand the simulation models adopted by Gamfeldt and Roger (2017) by combining the averaging and summing multifunctionality metrics together. We then propose a way to convert the two metrics into a scaling one, which makes multifunctionality comparable among studies. We also revisit the threshold-based approach to quantify multifunctionality (Byrnes, *et al.*, 2014b) and use the range-relevant standardization to deal with the mathematical artefact raised by Gamfeldt and Roger (2017). Finally, we compare biodiversity-multifunctionality relationships with the standardization methods using empirical studies from a number of ecosystems with an extensive range of climatic and edaphic conditions (Figure S1).

## METHODS

### A review of the averaging and summing approaches

Several quantitative approaches have been developed to examine biodiversity-multifunctionality relationships (Byrnes, *et al.*, 2014b, Dooley, *et al.*, 2015, Manning, *et al.*, 2018). Here, we reviewed two approaches. The averaging multifunctionality metric is one of the early and most commonly used approaches (Byrnes, *et al.*, 2014b, Hooper and Vitousek, 1998, Maestre, *et al.*, 2012, Mouillot *et al.*, 2011). A more recently developed approach is the summing multifunctionality metric (Meyer, *et al.*, 2018). Note that there are many versions of averaging indices that account for the standard deviation or geometric means of multiple single functions at the same time (see Table S1 for more information). For simplicity, we identify the two methods as the averaging approach and the summing approach.

The averaging or summing metric usually regresses with biodiversity, and the regression slopes are referred to as the effects of biodiversity on multifunctionality (Gamfeldt and Roger, 2017). One advantage of the averaging and summing approaches is that they are easy to calculate and interpret. Particularly, biodiversity-multifunctionality relationships can be analyzed using general linear models (Byrnes, *et al.*, 2014b, Gamfeldt and Roger, 2017, Maestre, *et al.*, 2012). However, the averaging approach has been criticized for a variety of statistical and biological reasons (Byrnes, *et al.*, 2014a, Dooley, 2018, Hines, 2019). For instance, the averaging metric has the same value when two functions have an identical value or when one function makes a higher contribution and another one makes a lower contribution (for example, the average of 5 and 5 is 5; the average of 1 and 9 is also 5) (Byrnes, *et al.*, 2014b, Dooley, 2018). In addition, although this limitation is equally true for the summing metric, we acknowledge that only one study focuses on the summing approach as it applies to the study of biodiversity-multifunctionality relationships (Meyer, *et al.*, 2018).



Simply put, this method has not been well explored, but we expect others will not adopt it without considering some of its limitations, such as weighting functions equally (but see Allan, *et al.*, 2015, Manning, *et al.*, 2018).

We explore the statistical underpinnings of the averaging and summing metrics using both simulation models and empirical data. We first note that the raw slope estimates (regression coefficients) do not help compare the strength of biodiversity effects when using these two multifunctionality metrics (Supplementary Note 1). In addition, we note that the averaging metric is mathematically equivalent to the summing metric if we standardize these two metrics to a common scale (*e.g.*, centering and scaling the two metrics into the scaling multifunctionality metric) (Figure 1; Supplementary Note 2). Most importantly, the values of the averaging and summing metrics are generally expressed in a relative scale. Therefore, we propose to use the scaling multifunctionality metric for data syntheses. Note that the scaling metric is not designed to solve the question of why biodiversity-multifunctionality relationships change with the number of functions considered. But this metric allows us to estimate the standardized slopes (or coefficients) of the general linear models between biodiversity and multifunctionality. As a result, the scaling approach could be used to compare results of different studies in a systematic way. Notably, there is a long history of using standardized coefficients in multiple regressions and structural equation models (Grace *et al.*, 2005, Grace *et al.*, 2018). Some of the commonly used standardization methods include Z-score and range-relevant standardization (Grace *et al.*, 2007, Grace and Bollen, 2005, Grace, *et al.*, 2018, Schielzeth, 2010). Here, we suggest a new approach – referred to as “scaling approach” – to convert the averaging or summing metrics into a single metric. In contrast to the traditional way, we use a Z-score transformation and then model the metric with the raw biodiversity data. Suppose that the estimated slope between the scaling metric and the raw biodiversity data is 0.2. We interpret this value to mean that, with each addition

of one unit of biodiversity (*i.e.*, an increase of one species), then multifunctionality increases by 0.2 standard deviations. Therefore, the semi-standardized slope estimate may provide an opportunity to make the multifunctionality metric comparable and interpretable across studies. Below we use simulation models to illustrate how different methods will affect our interpretation and comparison.

### **Simulation models for the averaging, summing and scaling approaches**

To our knowledge, Gamfeldt and Roger (2017) were the first to quantitatively and explicitly address whether biodiversity-multifunctionality relationships depend on the number of functions. Here, in contrast with their methods, we used three metrics (averaging, summing and scaling) to quantify multifunctionality. Note the summing metric is not the same as the one proposed by Meyer et al. (2018). They calculated a metric by summing the axis scores of the principal component analysis (PCA). Their metric has the advantage that it corrects for the correlations among individual ecosystem functions by calculating the components of PCA which are, by definition, uncorrelated with each other. In this study, we calculated the summing metric directly by summing individual standardized ecosystem functions without correcting for the correlation structure of ecosystem functions. The summing metric thus enables us to make direct comparisons with the averaging metric. In addition, the scaling metric is a modified metric that can be calculated either from the averaging or summing metrics (Figure 1). It is very similar but not exactly the same as Dooley's scaled average multifunctionality (SAM) metric (Dooley, 2018). However, the rationale for our metric is similar to SAM in that it converts the averaging or summing metric into a standardized one with a common scale (more information about the statistical rationales can be found in Supplementary Notes 1 and 2).

In brief, we simulated a species pool with 12 species, and each species performs nine ecosystem functions. Artificial communities were generated from the species pool. We simulated all possible combinations for the 12 species at each species richness level and generated a total of 4095 artificial communities at each step. We calculated ecosystem functions by averaging the function values that the species performed in the 4095 artificial communities. Meanwhile, we simulated a series of ecosystem functions subject to complementarity, which is one of the fundamental mechanisms to explain the observed positive effects of biodiversity on ecosystem functioning. We applied a saturating function for species richness and the complementarity factor (CF) to an ecosystem function in the following way:

$$CF = CF_{\max} \left( 1 - e^{1-S^r} \left( 1 - \frac{1}{CF_{\max}} \right) \right) \quad (\text{Eqn. 1})$$

where CF is one when species richness is one, and the maximum CF ( $CF_{\max}$ ) is three.  $S$  is species richness and  $r$  is the rate of CF reaching its maximum value ( $CF_{\max}$ ). Each single function was standardized either with Z-score or its maximum value. In total, we simulated nine scenarios and generated various sets of functions that varied by the number functions subject to complementarity. All possible combinations of the nine single ecosystem functions were used to calculate the averaging, summing or scaling multifunctionality metrics in each scenario. We refer to Gamfeldt and Roger (2017) for more details of the simulations.

## A review of the threshold-based approach

A second common approach to quantify multifunctionality is the threshold-based approach (Byrnes, *et al.*, 2014b, Gamfeldt, *et al.*, 2008, Zavaleta, *et al.*, 2010). Generally, it converts the matrix of ecosystem functions ( $n$  communities by  $m$  functions) into a binary matrix for a given threshold (or cut-off) and counts the number of functions that surpass a given threshold in a community. This process is very similar to the method used to calculate species richness in a community where species abundance is converted into a binary species table (presence and absence matrix) and the number of species in the community is counted. Next, the number of counted functions is fitted to a generalized linear model against biodiversity, and the slope of this linear model is referred as the effects of biodiversity on multifunctionality. The advantage of the threshold-based approach is that it can identify whether two functions reach to high levels of functionality at the same time. However, the biological interpretation for the threshold-based approach remains challenging (Gamfeldt and Roger, 2017). In addition, the threshold-based approach is sensitive to the thresholds used to count the number of functions and the methods used to standardize single ecosystem function (Byrnes, *et al.*, 2014b, Gamfeldt and Roger, 2017). Most importantly, Gamfeldt and Roger (2017) pointed out that both positive and negative effects of biodiversity on multifunctionality change with the number of functions, and that this pattern is likely due to a mathematical artefact (Supplementary Note 3). Therefore, the threshold-based approach may inhibit reliable comparisons across studies.

As we discuss above, the range-relevant standardization (Grace and Bollen, 2005, Grace, *et al.*, 2018) may solve the mathematical issue and make the estimated slopes comparable across studies. Generally, the range of the number of functions surpassing a given threshold varies from zero to the total number of functions measured. Thus, if we standardize the

number of functions surpassing a given threshold, we will obtain a series of standardized values ranging from zero to one. The slope estimates between the standardized number of functions and biodiversity should be comparable to a common scale.

To support our speculation, we used soil microbial biomass from a pool of 82 ecosystem function indicators measured in an empirical study at the Jena biodiversity experiment (Meyer, *et al.*, 2018, Weisser *et al.*, 2017). Similar to Gamfeldt and Roger (2017), we simulated a variety of scenarios by replicating soil microbial biomass many times (*i.e.*, 5, 15, 25, 35 and 45 number of functions in total considered). Note that the simulations focused on a single ecosystem function, but we treated the replications of microbial biomass as multiple ecosystem functions. Such scenarios are not ecologically relevant in a real study of biodiversity-multifunctionality relationships, but they illustrate the rationale and the cause of the mathematical artefact of the threshold-based approach. In addition, we used a hierarchical sampling approach, by which we randomly sampled a subset of the 82 functions from the Jena dataset (Supplementary Note 3). The hierarchical sampling approach enabled us to explore whether biodiversity-multifunctionality relationships changed when novel functions are included in estimate of multifunctionality.

### **Empirical data acquisition and description**

We compiled five empirical datasets including two manipulative biodiversity experiments across nine grasslands and three observational studies spanning 300 sites around the world (Figure S1; Table S2). The two manipulative biodiversity studies include one grassland biodiversity experiment that measured 82 ecosystem functions (Jena grassland) (Meyer, *et al.*, 2018, Weisser, *et al.*, 2017) and eight pan-European BIODDEPTH grassland biodiversity experiments that measured six functions (Spehn *et al.*, 2005). The three observational studies include 208 forest plots of the FunDivEUROPE platform with 26 measured functions in each

single plot (Ratcliffe *et al.*, 2017), 236 global drylands with 14 measured functions (Maestre, *et al.*, 2012, Ochoa-Hueso *et al.*, 2018) and 60 Tibetan grasslands with eight measured functions (Jing *et al.*, 2015). These five studies span local, regional and global spatial scales with diverse climates, vegetation and soil types. The ecosystem functions compiled from each study represent typically measured functions in terrestrial ecosystems (*e.g.*, primary productivity, nitrogen and phosphorus pools in plants and soils, soil microbial biomass and enzymatic activities). All the surrogates of ecosystem functions used are either direct measures of ecosystem stocks (*e.g.*, plant biomass, soil C stock), fluxes (*e.g.*, productivity, decomposition) or indirect measures of ecosystem properties (*e.g.*, light interception) (Meyer, *et al.*, 2018). We focused on plant species richness because this was the only biodiversity metric common to all datasets.

### **Statistical analyses**

To make our simulations comparable to Gamfeldt and Roger (2017) and to estimate biodiversity effects on multifunctionality, we used two data transformations – standardized to maximum value and Z-score – for the single ecosystem functions prior to the estimate of multifunctionality. For the observational studies and biodiversity experiments, we used general linear models without controlling for other abiotic variables (*e.g.*, climate, land use intensity, and soil properties). Although these abiotic variables may affect the strength of biodiversity-multifunctionality relationships (Duffy *et al.*, 2017), they will not affect our interpretation of the effects of biodiversity on multifunctionality.

To estimate the average effect of biodiversity on multifunctionality, we fitted linear regression models with all combinations of functions from one to the maximum number of functions considered within each simulation or empirical dataset; thus our method combined two multifunctionality approaches, the single functions approach (Byrnes, *et al.*, 2014b) and

the multifunctionality metric approach. For three datasets – Jena grassland, European forests and Global drylands (Maestre, *et al.*, 2012, Meyer, *et al.*, 2018, Ratcliffe, *et al.*, 2017) – if the ecosystem function combinations were greater than 500, we randomly sampled 500 combinations of functions and calculated multifunctionality for each combination of those single functions. We extracted the slopes and standard errors (s.e.) from linear models. We calculated the 95% confidence interval ( $CI = 1.96 \times \text{s.e.}$ ) for each slope estimate. If the lower 95% CI was higher than zero or the upper 95% CI was lower than zero, we categorized the slope estimates as positive or negative effects of biodiversity on multifunctionality, respectively.

Using the threshold-based approach we further compared the effects of biodiversity on multifunctionality using the five empirical datasets. Specifically, we followed the approach adopted by Byrnes *et al.* (2014b). We first standardized single functions to their maximum values by taking the mean of the six highest measures of each function as the maximum value. We counted the number of functions surpassing a given threshold ranging from 0.05 to 0.99. We conducted the threshold-based analysis by fitting the generalized linear models with the raw number of functions and the range standardized number of functions upon biodiversity.

All statistical analyses were conducted in R version 3.3.3 (R Development Core Team, 2019). The threshold approach was conducted using the package ‘multifunc’ (Byrnes, *et al.*, 2014b).

## RESULTS

### **Biodiversity-multifunctionality relationships vary with the averaging, summing and scaling approaches**

Our simulation models revealed that the strength of biodiversity-multifunctionality relationships varied with the approaches used to estimate multifunctionality (Figure 2). Generally, the averaging approach showed that the average estimated slopes relating biodiversity to multifunctionality did not change as the number of functions increased for a given number of functions subject to complementarity (*e.g.*, three out of nine functions subject to complementarity). However, the strength of the average estimated slopes increased as the number of functions subject to complementarity increased (from zero to nine). The summing and scaling approaches demonstrated that the average estimated slopes increased with the number of functions. Furthermore, we found that the averaging and summing approaches were sensitive to the methods used to standardize single function (standardized to maximum value vs. Z-score). That is, the strength of biodiversity-multifunctionality relationships estimated by the Z-score for single function was ~ ten times higher than the maximum standardization (Figure 2 and Figure S2). In contrast, we obtained the same biodiversity-multifunctionality relationships using the scaling approach whichever the standardization methods used (Figure 2 and Figure S2). Finally, when we grouped biodiversity-multifunctionality relationships into two categories (significantly positive and neutral), we found that, when functions were subject to complementarity, the positive effects of biodiversity on multifunctionality were most commonly observed when the number of functions included increased (Supplementary Note 4 and Figure S3).

We found qualitatively similar results when we examined the empirical datasets. The average effects of biodiversity on multifunctionality did not change with the number of functions



included in the averaging multifunctionality metric, but it increased with the summing and scaling multifunctionality metrics (Figure 3). Results of biodiversity experiments such as the Jena grassland and European grasslands, were similar with those of observational studies such as European forests, Tibetan grasslands and global drylands. Among all the studies, the European forests had the largest variability in the effects of biodiversity on single ecosystem functions (Figure 4a) and scaling multifunctionality (Figure 4b). Biodiversity-multifunctionality relationships in temperate grasslands were as strong as they were in global drylands (Figure 4b). Furthermore, the probability of biodiversity having positive, neutral or negative effects on multifunctionality changed with the number of functions included (Supplementary Note 4 and Figure S4). Positive biodiversity effects on multifunctionality were more common as the number of functions included increased in the Jena grassland, European grasslands, Tibetan grasslands and global drylands. In contrast, neutral biodiversity effects on multifunctionality were more common as the number of functions included increased in the European forests.

### **Biodiversity-multifunctionality relationships vary with the standardization methods of the threshold-based approach**

For the simulations using soil microbial biomass data from Jena grassland, we found that the effects of biodiversity on multifunctionality (raw slope estimates) increased as the number of functions increased (see Supplementary Note 3 for details). However, the effects of biodiversity on multifunctionality was identical at a given threshold when the standardized slopes were estimated. When exploring the other 82 functions measured in the Jena experiment, we found similar results to what we found with soil microbial biomass (data not shown). In addition, using the hierarchical sampling simulations, we found that increasing the number of functions increased the strength of biodiversity-multifunctionality relationships

when we used the raw number of functions that surpassed a given threshold (Supplementary Note 3). In contrast, we found that the average effect of biodiversity on multifunctionality did not change with the number of functions in total considered when we standardized the number of function surpassing a given threshold. Introducing novel functions for the estimate of multifunctionality decreased the range of 95% confidence intervals.

For the empirical datasets, we compared the results using the raw number of functions and the range-relevant standardized number of functions. The trends using the raw number of functions were similar with those using the range-relevant standardized number of functions, but they were different in the strength of biodiversity effects (Figure 5). That is, the effects of biodiversity on multifunctionality increased with the number of functions included using the raw number of functions. Using the range-relevant standardized number of functions, our results showed that biodiversity in the European forests had the largest effects and highest variability on multifunctionality, and that biodiversity in the Tibetan grasslands also had larger effects than did biodiversity in the Jena grassland, European grasslands and global drylands (Figure 5). Most importantly, the results of the threshold-based approach using the range standardized number of functions (Figure 5) were generally in line with the results of the scaling approach (Figure 4).

## DISCUSSION

In this study, we demonstrated that the effects of biodiversity on multifunctionality varied across simulations and empirical datasets. Most interestingly, using the same dataset but different multifunctionality metrics, the effects of biodiversity on multifunctionality either increased or did not change with the number of functions. Although some previous studies found similar results when applying the averaging and summing multifunctionality metrics and the threshold-based approach (Fanin, *et al.*, 2018, Gamfeldt and Roger, 2017, Meyer, *et al.*, 2018), we note that these results depended primarily on the underlying statistics of these metrics rather than “real” effects of biodiversity on multifunctionality. That is, the result arises because of math, not ecology. Below, we firstly provide our statistical explanations for whether biodiversity-multifunctionality relationships change with the number of functions. Then we discuss how the standardization methods can be used to improve our understanding of biodiversity-multifunctionality relationships across studies.

First, for the averaging approach, biodiversity-multifunctionality relationships are obtained from the relationship between biodiversity and single ecosystem function as follows (see proof of the equation in Supplementary Note 2):

$$\beta_{EMFav} = \frac{1}{n} \sum_{i=1}^n \beta_i \quad (\text{Eqn. 2})$$

where  $\beta_{EMFav}$  is the slope estimate between biodiversity and multifunctionality. It can be expressed as the expected average effect of biodiversity on multifunctionality.  $\beta_i$  is the slope estimate of any one of the regression slopes ranging from one to  $n$  single ecosystem functions. In an extreme case in which these single functions have identical slope estimates for the relationships between biodiversity and single ecosystem functions, all the  $n$  single

functions would have the same regression slopes  $\beta_1 = \beta_2 = \dots = \beta_n$ , then we would obtain the following regression slopes for the averaging approach:

$$\beta_{EMFav} = \frac{1}{n} \sum_{i=1}^n \beta_i = \frac{1}{n} (n \times \beta_i) = \beta_i \quad (\text{Eqn. 3})$$

The effects of biodiversity on multifunctionality ultimately equal the average slope estimates between biodiversity and single ecosystem functions, and thus biodiversity-multifunctionality relationships would not change as the number of functions increased. Furthermore, the population mean may become higher when there are more ecosystem functions subject to complementarity because of higher slope coefficients obtained. Therefore, we can explain why the strength of the average biodiversity effect on multifunctionality increases with the number of functions subject to complementarity (Gamfeldt and Roger, 2017). In addition, since the population mean increases with the number of functions subject to complementarity, the fraction of biodiversity having a positive effect on multifunctionality increases with the number of functions subject to complementarity in the same way (Figures S3 and S4).

Second, for the summing approach, biodiversity-multifunctionality relationships can be obtained from the relationship between biodiversity and single ecosystem functions as follows:

$$\beta_{EMFsum} = \sum_{i=1}^n \beta_i \quad (\text{Eqn. 4})$$

In an extreme case the same as the averaging approach in which these single functions have identical slope estimates, we would obtain the following regression slopes for the summing approach:

$$\beta_{EMFsum} = \sum_{i=1}^n \beta_i = n \times \beta_i \quad (\text{Eqn. 5})$$

The expected effects of biodiversity on the summing multifunctionality tend to approach the sum of the slope estimates between biodiversity and single ecosystem functions. Thus, the biodiversity-multifunctionality relationships would increase with the number of ecosystem functions. In comparison to the averaging approach, our results showed that the summing approach gave different weights to the importance of single ecosystem functions (Höting, *et al.*, 2019). Consider the hypothetical example that we have three ecosystem functions and the effects of biodiversity on each single ecosystem functions are 0.2, 0.4 and 0.6, respectively. The averaging approach would give a weight of 1/3 to each function while the summing approach would give a weight of 1. Therefore, we would derive the effects of biodiversity on multifunctionality to be 0.5 for the averaging approach and 1.2 for the summing approach. Therefore, the underlying assumptions are that for the averaging approach biodiversity affects the level of multifunctionality via averaging effects of single functions, while for the summing approach biodiversity affects multifunctionality via simple additive effects of single functions.

Third, for both the averaging and summing approaches, the methods used to standardize single function influenced the strength of biodiversity-multifunctionality relationships (Figures 1 and S2). However, the scaling multifunctionality metric was not sensitive to the methods used for standardizing single ecosystem functions (Figures 1 and S2). The scaling metric thus puts the slope estimates into a common scale and aides in comparisons among different empirical studies. Traditionally, the standardized coefficients (or slope estimates) can be obtained in two ways (Grace, *et al.*, 2018). The first is to standardize the response variables and predictors using Z-score or range-relevant standardization before fitting the models. The second is to obtain the raw slope estimates in advance and then standardize the

slope estimates by the ratios of the standard deviations of predictors and response variables. In this study, we used the former way to derive the standardized slope coefficients. One of the advantages of standardized coefficients are to provide a common scale to compare the relative importance of predictors (*e.g.*, biodiversity, climate and soil physicochemical properties) on response variables for between-study comparisons (Schielzeth, 2010). The idea to use the scaling metric was consistent with the methods often used in the study of biodiversity-multifunctionality relationships, in particular estimating the standardized coefficients by using either structural equation models (Delgado-Baquerizo *et al.*, 2016b, Jing, *et al.*, 2015, Lefcheck *et al.*, 2015b, Wang *et al.*, 2019) or bivariate and multivariate linear models (Le Bagousse-Pinguet *et al.*, 2019, Soliveres *et al.*, 2016a, Soliveres *et al.*, 2016b). Therefore, we argue that this approach, which could provide a standard way to compare results among studies, should become the standard approach for synthetic studies or for studies that compare the influence of biodiversity on multifunctionality. However, we caution that the influence of increasing biodiversity on the scaling multifunctionality metric ( $\beta_{EMF_{Sca}}$ ) with the number of functions does not imply a biological mechanism that drives biodiversity-multifunctionality relationships (Supplementary Note 2). The increase in  $\beta_{EMF_{Sca}}$  could arise from the decline in the variation in the level of multifunctionality and the average pairwise correlations among single functions with increasing number of functions considered (Supplementary Note 2).

Finally, for the threshold-based approach, when increasing the raw number of functions, a mathematical artefact emerged as increased the effects of biodiversity on multifunctionality. Gamfeldt and Roger (2017) also found that when the number of functions in total considered varies, the threshold-based approach might be not suitable for comparing the effects of biodiversity on multifunctionality among studies. The mathematical artefact appeared to contradict the original idea of the threshold-based approach (Byrnes, *et al.*, 2014b, Lefcheck,

*et al.*, 2015a). However, we demonstrated that this mathematical artefact was due to the upper limits of the number of functions per addition of one species increasing with the number of functions in total considered (Supplementary Note 3). When we used the range-relevant standardization, we found that the standardized slope estimates between biodiversity and the number of functions surpassing a given threshold were identical for soil microbial biomass in the Jena biodiversity experiment. This finding was in line with a recent analysis of the Global drylands dataset (Le Bagousse-Pinguet, *et al.*, 2019), which the biodiversity effects on multifunctionality were comparable when the standardized regression coefficients were used to quantify the effects of biodiversity on multifunctionality.

Over the past decade, we have moved toward a better understanding of how biodiversity is related to multiple ecosystem functions simultaneously. However, in this study, we showed that the interpretation of biodiversity-multifunctionality relationships varies with the method considered, and that the standardization methods provide a valuable opportunity to improve our understanding of the statistical mechanisms underlying multifunctionality. The choice of standardization methods is thus critical and has a significant impact on the outcomes of a study (Gamfeldt and Roger, 2017, Schmid *et al.*, 2017). However, to date, there has not been a unified or best method for estimating ecosystem multifunctionality (Byrnes, *et al.*, 2014b, Hölting, *et al.*, 2019, Manning, *et al.*, 2018), and different approaches for quantifying biodiversity-multifunctionality relationships yield different results (Gamfeldt and Roger, 2017, Meyer, *et al.*, 2018); this is an obvious and dispiriting set of circumstances if the goal of a research is to advance our understanding of how biodiversity affects ecosystem multifunctionality.

The standardized multifunctionality metrics we propose here are desirable and have several advantages when examining spatial or temporal variation in biodiversity-multifunctionality

relationships. First, they enable the interpretation of biodiversity effects easily. Second, they facilitate between-study comparisons by estimating a standardized effect size for quantitative reviews and meta-analyses. We therefore suggest the use of standardized multifunctionality metrics (*e.g.*, scaling or range-relevant metrics), especially when the studies use different numbers of ecosystem functions or measure different ecosystem functions. There are, however, three important caveats. First, standardized multifunctionality metrics cannot overcome the limitations of the averaging approach, threshold-based approach or the other commonly used approaches (Byrnes, *et al.*, 2014b, Dooley, *et al.*, 2015). Second, different ecosystem properties (and hence multifunctionality metrics) might be valued in different ecological contexts, so some researchers may prefer other multifunctionality metrics (Hölting, *et al.*, 2019). For example, when researchers quantify multifunctionality, the idea is not necessarily to make a metric that is comparable across studies, instead they might be interested in optimizing overall ecosystem functioning by exploring different scenarios of stakeholder priorities given to single ecosystem functions (Allan, *et al.*, 2015, Manning, *et al.*, 2018, Slade, *et al.*, 2017). Finally, when individual functions are driven by different factors (Bradford *et al.*, 2014b), it would be interesting to explore the effects of biodiversity by using single functions approach and multiple indices of overall ecosystem functioning (Byrnes, *et al.*, 2014b, Ratcliffe, *et al.*, 2017).

In this study, we have focused on plant species richness as the measure of biodiversity in part because this is easier than trying to estimate abundance, and in part because most early experimental studies focused on biodiversity as the number of species and controlled the number of species in plots. However, in natural communities, or in hyper-diverse communities, it is challenging to detect all of the species present (Chao *et al.*, 2017). It would be interesting to use abundance-based diversity indices in future studies such as the exponential of Shannon entropy and the inverse of Simpson index. Those indices are the



effective numbers of species and estimate the true diversity by giving different weights to rare vs. common species (Jost, 2006). In addition, our simulations and statistical analyses focused on multifunctionality metrics with Gaussian distributions. We did not evaluate whether other data distributions (*e.g.*, binary, log-normal and skewed) influenced the metric performance (Schoolmaster *et al.*, 2012). When there is a combination of distributions in the data, the convergence of statistical models may disappear because of deviations in normality and linearity. The range-relevant standardization (Grace, *et al.*, 2018) might be more appropriate in this situation. Future studies should also focus on deepening our understanding of how ecosystem functions interact with each other (*e.g.*, trade-offs and synergies) (Dooley, *et al.*, 2015, Meyer, *et al.*, 2018, Slade *et al.*, 2019) and the underlying statistical underpinnings of different multifunctionality metrics (Hölting, *et al.*, 2019). This would improve the interpretation and comparison of biodiversity-multifunctionality relationships.

## CONCLUSION

The most interesting result from our work is that we can now compare, on a level playing field, how biodiversity is related to multifunctionality across studies that differ in scale and system. Similarly with van der Plas (2019), we found that biodiversity is key driver of multifunctionality in naturally and randomly assembled communities. The scaling and range-relevant approaches allow us to do this because they put biodiversity-multifunctionality relationships on a common scale. Our approach would allow comparisons among systems that differ in land-use history, climate, soil properties, or other factors. In fact, it will increase the likelihood that we understand how those factors might alter the relationship between biodiversity and multifunctionality in a changing world. Taken together, our findings suggest that standardization methods are a useful way of comparing biodiversity-multifunctionality relationships among studies.

## FUNDING

This work was supported by the National Natural Science Foundation of China (31600428) to X.J. and a Semper Ardens grant from Carlsberg Foundation to N.J.S.. F.T.M. and the global drylands dataset were supported by the European Research Council (ERC Grant Agreements 242658 [BIOCOM] and 647038 [BIODESERT]).

## ACKNOWLEDGEMENTS

A sincere thank you to Mingzhen Lu and Wenting Feng for their diligent proofreading of this paper. We thank Eric Allan, Forest Isbell, members of the Classen laboratory, the handling editor and two anonymous reviewers for providing insightful comments.

## CODE AND DATA AVAILABILITY

R scripts are deposited to Github (<https://github.com/XJingPKU/BEMFmetric>). Part of the empirical datasets from the Tibetan grasslands are deposited to Zenodo (<https://zenodo.org/record/2650366#.XMCxZDBKjb0>). The other datasets can be found within each of the published papers (Maestre, *et al.*, 2012, Meyer, *et al.*, 2018, Ochoa-Hueso, *et al.*, 2018, Ratcliffe, *et al.*, 2017, Spehn, *et al.*, 2005).

## REFERENCES

- Allan E, Manning P, Alt F, *et al.* (2015). Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecol. Lett.* **18**:834-43.
- Bradford MA, Wood SA, Bardgett RD, *et al.* (2014a). Reply to byrnes et al.: Aggregation can obscure understanding of ecosystem multifunctionality. *Proc. Natl. Acad. Sci. USA* **115**:E5491.
- Bradford MA, Wood SA, Bardgett RD, *et al.* (2014b). Discontinuity in the responses of ecosystem processes and multifunctionality to altered soil community composition. *Proc. Natl. Acad. Sci. USA* **111**:14478-83.
- Butterfield BJ, Camhi AL, Rubin RL, *et al.* (2016). Tradeoffs and compatibilities among ecosystem services: Biological, physical and economic drivers of multifunctionality. *Adv. Ecol. Res.* **54**:207-43.
- Byrnes J, Lefcheck JS, Gamfeldt L, *et al.* (2014a). Multifunctionality does not imply that all functions are positively correlated. *Proc. Natl. Acad. Sci. USA* **111**:E5490.
- Byrnes JEK, Gamfeldt L, Isbell F, *et al.* (2014b). Investigating the relationship between biodiversity and ecosystem multifunctionality: Challenges and solutions. *Methods Ecol. Evol.* **5**:111-24.
- Chao A, Chiu CH, Colwell RK, *et al.* (2017). Deciphering the enigma of undetected species, phylogenetic, and functional diversity based on good - turing theory. *Ecology* **98**:2914-29.
- Delgado-Baquerizo M, Giaramida L, Reich PB, *et al.* (2016a). Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *J. Ecol.* **104**:936-46.
- Delgado-Baquerizo M, Maestre FT, Reich PB, *et al.* (2016b). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **7**:10541.
- Dooley A (2018). Modelling techniques for biodiversity and ecosystem multifunctionality-theoretical development and application, National University of Ireland Maynooth.

- Dooley A, Isbell F, Kirwan L, *et al.* (2015). Testing the effects of diversity on ecosystem multifunctionality using a multivariate model. *Ecol. Lett.* **18**:1242-51.
- Duffy JE, Godwin CM, Cardinale BJ (2017). Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature* **549**:261-4.
- Duffy JE, Richardson JP, Elizabeth AC (2003). Grazer diversity effects on ecosystem functioning in seagrass beds. *Ecol. Lett.* **6**:637-45.
- Fanin N, Gundale MJ, Farrell M, *et al.* (2018). Consistent effects of biodiversity loss on multifunctionality across contrasting ecosystems. *Nat. Ecol. Evol.* **2**:269-78.
- Gamfeldt L, Hillebrand H, Jonsson PR (2008). Multiple functions increase the importance of biodiversity for overall ecosystem functioning. *Ecology* **89**:1223-31.
- Gamfeldt L, Roger F (2017). Revisiting the biodiversity-ecosystem multifunctionality relationship. *Nat. Ecol. Evol.* **1**:0168.
- Gotelli NJ, Ulrich W, Maestre FT (2011). Randomization tests for quantifying species importance to ecosystem function. *Methods Ecol. Evol.* **2**:634-42.
- Grace JB, Anderson T, Smith MD, *et al.* (2007). Does species diversity limit productivity in natural grassland communities? *Ecol. Lett.* **10**:680-9.
- Grace JB, Bollen KA (2005). Interpreting the results from multiple regression and structural equation models. *Bull. Ecol. Soc. Am.* **86**:283-95.
- Grace JB, Johnson DJ, Lefcheck JS, *et al.* (2018). Quantifying relative importance: Computing standardized effects in models with binary outcomes. *Ecosphere* **9**:e02283.
- Hautier Y, Isbell F, Borer ET, *et al.* (2018). Local loss and spatial homogenization of plant diversity reduce ecosystem multifunctionality. *Nat. Ecol. Evol.* **2**:50-6.

Hector A, Bagchi R (2007). Biodiversity and ecosystem multifunctionality. *Nature* **448**:188-90.

Hector A, Schmid B, Beierkuhnlein C, *et al.* (1999). Plant diversity and productivity experiments in european grasslands. *Science* **286**:1123-7.

Hines J (2019). Ecosystem functioning: How much system is needed to explain function? *Curr. Biol.* **29**:R1072-R4.

Hölting L, Beckmann M, Volk M, *et al.* (2019). Multifunctionality assessments – more than assessing multiple ecosystem functions and services? A quantitative literature review. *Ecol. Indicators* **103**:226-35.

Hooper DU, Adair EC, Cardinale BJ, *et al.* (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486**:105-8.

Hooper DU, Vitousek PM (1998). Effects of plant composition and diversity on nutrient cycling. *Ecol. Monogr.* **68**:121-49.

Isbell F, Calcagno V, Hector A, *et al.* (2011). High plant diversity is needed to maintain ecosystem services. *Nature* **477**:199-202.

Jing X, Sanders NJ, Shi Y, *et al.* (2015). The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nat. Commun.* **6**:8159.

Jost L (2006). Entropy and diversity. *Oikos* **113**:363-75.

Le Bagousse-Pinguet Y, Soliveres S, Gross N, *et al.* (2019). Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality. *Proc. Natl. Acad. Sci. USA* **116**:8419-24.

Lefcheck JS, Byrnes JE, Isbell F, *et al.* (2015a). Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat. Commun.* **6**:6936.

- Lefcheck JS, Duffy JE (2015b). Multitrophic functional diversity predicts ecosystem functioning in experimental assemblages of estuarine consumers. *Ecology* **96**:2973-83.
- Maestre FT, Quero JL, Gotelli NJ, *et al.* (2012). Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335**:214-8.
- Manning P, van der Plas F, Soliveres S, *et al.* (2018). Redefining ecosystem multifunctionality. *Nat. Ecol. Evol.* **2**:427-36.
- Meyer ST, Ptacnik R, Hillebrand H, *et al.* (2018). Biodiversity-multifunctionality relationships depend on identity and number of measured functions. *Nat. Ecol. Evol.* **2**:44-9.
- Miki T, Yokokawa T, Matsui K (2014). Biodiversity and multifunctionality in a microbial community: A novel theoretical approach to quantify functional redundancy. *P. Roy. Soc. B-Biol. Sci.* **281**:20132498.
- Mori AS, Isbell F, Fujii S, *et al.* (2016). Low multifunctional redundancy of soil fungal diversity at multiple scales. *Ecol. Lett.* **19**:249-59.
- Mouillot D, Villegger S, Scherer-Lorenzen M, *et al.* (2011). Functional structure of biological communities predicts ecosystem multifunctionality. *PLoS One* **6**:e17476.
- Ochoa-Hueso R, Eldridge DJ, Delgado-Baquerizo M, *et al.* (2018). Soil fungal abundance and plant functional traits drive fertile island formation in global drylands. *J. Ecol.* **106**:242-53.
- R Development Core Team (2019). R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria. <https://www.R-project.Org/>.
- Ratcliffe S, Wirth C, Jucker T, *et al.* (2017). Biodiversity and ecosystem functioning relations in european forests depend on environmental context. *Ecol. Lett.* **20**:1414-26.
- Schiele H (2010). Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**:103-13.

Schmid B, Baruffol M, Wang Z, *et al.* (2017). A guide to analyzing biodiversity experiments. *J. Plant Ecol.* **10**:91-110.

Schoolmaster DR, Grace JB, Schweiger EW (2012). A general theory of multimetric indices and their properties. *Methods Ecol. Evol.* **3**:773-81.

Slade EM, Bagchi R, Keller N, *et al.* (2019). When do more species maximize more ecosystem services? *Trends Plant Sci.* **24**:790-3.

Slade EM, Kirwan L, Bell T, *et al.* (2017). The importance of species identity and interactions for multifunctionality depends on how ecosystem functions are valued. *Ecology* **98**:2626-39.

Snelgrove PV, Thrush SF, Wall DH, *et al.* (2014). Real world biodiversity-ecosystem functioning: A seafloor perspective. *Trends Ecol. Evol.* **29**:398-405.

Soliveres S, Manning P, Prati D, *et al.* (2016a). Locally rare species influence grassland ecosystem multifunctionality. *Phil. Trans. R. Soc. B* **371**:20150269.

Soliveres S, van der Plas F, Manning P, *et al.* (2016b). Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature* **536**:456-9.

Song M-H, Zhu J-F, Li Y-K, *et al.* (2020). Shifts in functional compositions predict desired multifunctionality along fragmentation intensities in an alpine grassland. *Ecol. Indicators* **112**:106095.

Spehn E, Hector A, Joshi J, *et al.* (2005). Ecosystem effects of biodiversity manipulations in european grasslands. *Ecol. Monogr.* **75**:37-63.

Tilman D, Isbell F, Cowles JM (2014). Biodiversity and ecosystem functioning. *Annu. Rev. Ecol. Evol. Syst.* **45**:471-93.

van der Plas F (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biol. Rev. Camb. Philos. Soc.* **94**:1220-45.

van der Plas F, Manning P, Allan E, *et al.* (2016). Jack-of-all-trades effects drive biodiversity-ecosystem multifunctionality relationships in european forests. *Nat. Commun.* **7**:11109.

Wagg C, Schlaeppli K, Banerjee S, *et al.* (2019). Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nat. Commun.* **10**:1-10.

Wang L, Delgado-Baquerizo M, Wang D, *et al.* (2019). Diversifying livestock promotes multidiversity and multifunctionality in managed grasslands. *Proc. Natl. Acad. Sci. USA* **116**:6187-92.

Weisser WW, Roscher C, Meyer ST, *et al.* (2017). Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: Patterns, mechanisms, and open questions. *Basic Appl. Ecol.* **23**:1-73.

Zavaleta ES, Pasari JR, Hulvey KB, *et al.* (2010). Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proc. Natl. Acad. Sci. USA* **107**:1443-6.

Accepted Manuscript



## FIGURE CAPTIONS

**Figure 1.** An example shows the effects of biodiversity on single functions and ecosystem multifunctionality (EMF). We simulate two ecosystem functions (F1 and F2), each with a standardized normal distribution. Numbers in the bars are slope estimates. The blue lines of the inset show the slopes of linear regression between biodiversity and ecosystem functions. The orange lines in the middle of the inset indicate the slopes obtained either from the averaging metric or from the summing metric. As illustrated in the figure, the Z score-transformed averaging and summing metrics have the same slope estimate.

**Figure 2.** Summary of the simulations shows the effects of biodiversity on multifunctionality (EMF) in related to the number of functions included in the EMF estimate. The jittered points are the slope estimates showing the effects of biodiversity on EMF in each combination of the nine functions considered. Lines (average biodiversity effect on EMF) are the slopes of linear regression between the biodiversity-multifunctionality (all combinations of the nine functions) relationships and the number of ecosystem functions. Note that we use different ranges of y-axis to aid in visualization.

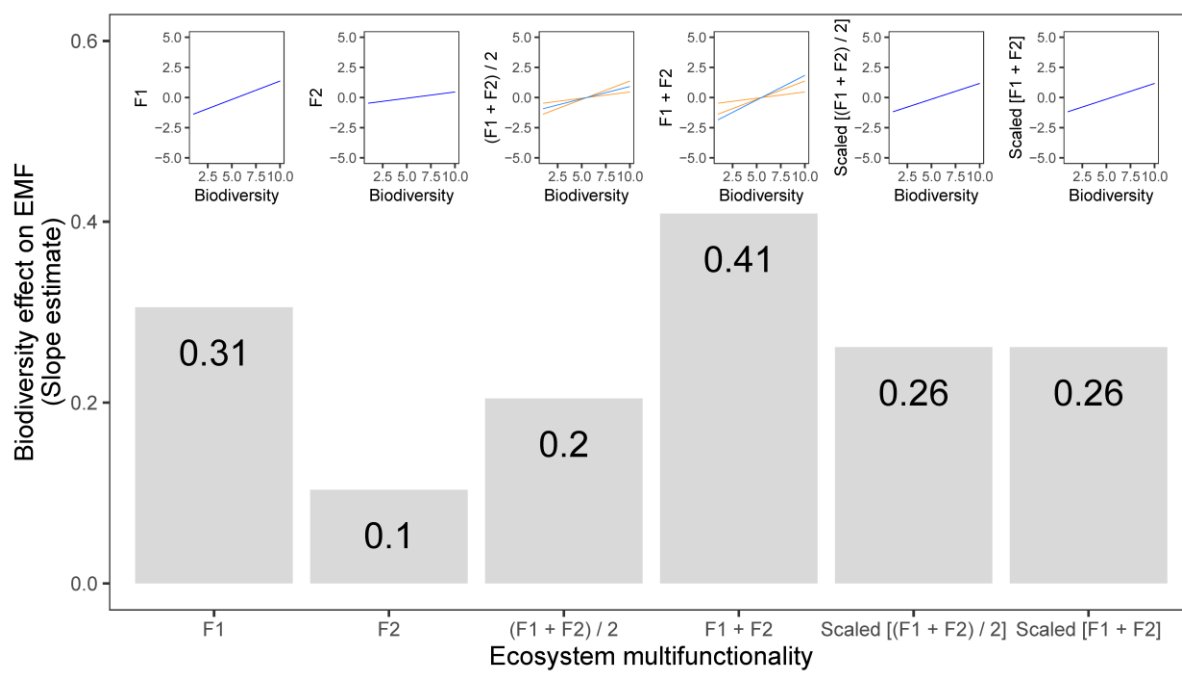
**Figure 3.** Relationships between the number of functions included in the estimate of multifunctionality (EMF) and the effects of biodiversity on EMF in five empirical datasets. Black lines (average biodiversity effect on EMF) are the slopes of linear regression between the number of functions and the biodiversity-EMF relationships. Red points indicate significantly positive effects of biodiversity on EMF, gray points indicate neutral effects, and blue points indicate significantly negative effects of biodiversity.

**Figure 4.** Effects of biodiversity on single ecosystem functions and multifunctionality (EMF) across five empirical datasets. (a) Each boxplot shows the effects of biodiversity on single ecosystem functions; each jittered point represents the slope estimate of the relationship between biodiversity and the single ecosystem functions; the lines in the box denote median, the middle 50% of the data are represented by the interquartile range box and the bottom 25% and the top 25% of the data are represented by the whiskers. (b) The density plot shows the effects of biodiversity on EMF (the scaling metric is reported here).

**Figure 5.** Effects of biodiversity on multifunctionality using the threshold-based approach across five empirical datasets. The upper panels summarize the results using unstandardized number of functions surpassing a given threshold. The lower panels summarize the results using standardized number of functions, which puts the effect size (change in number of functions per addition of one species) on a common and comparable scale across different studies.

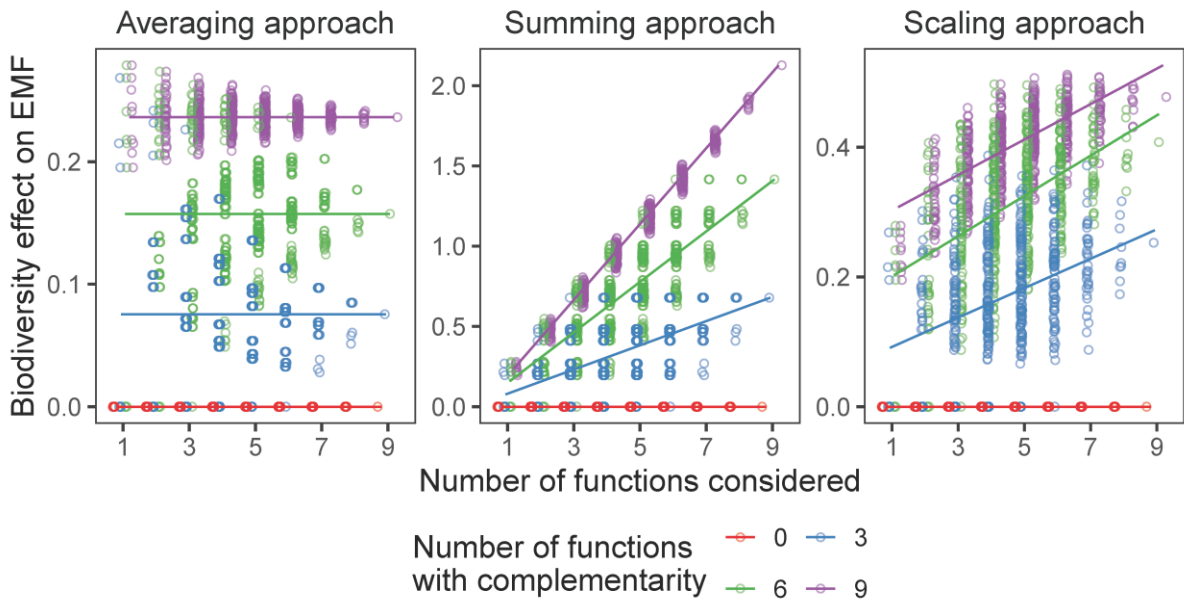
Accepted Manuscript

Figure 1



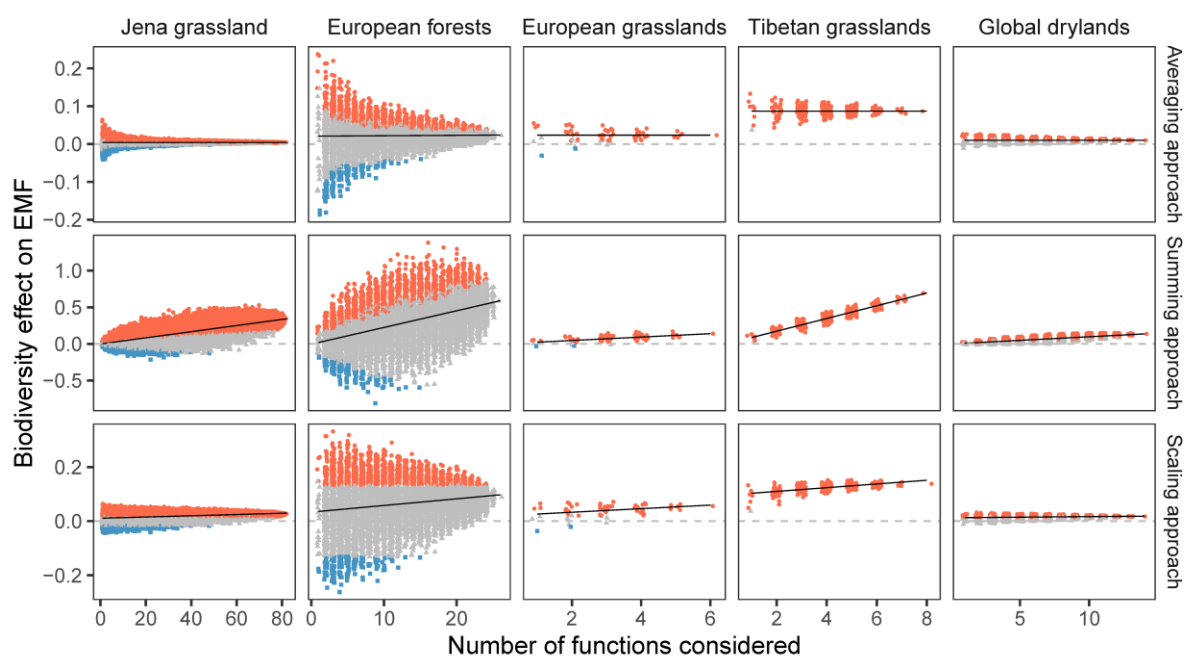
Accepted Manuscript

Figure 2



Accepted Manuscript

Figure 3



Accepted Manuscript

Figure 4

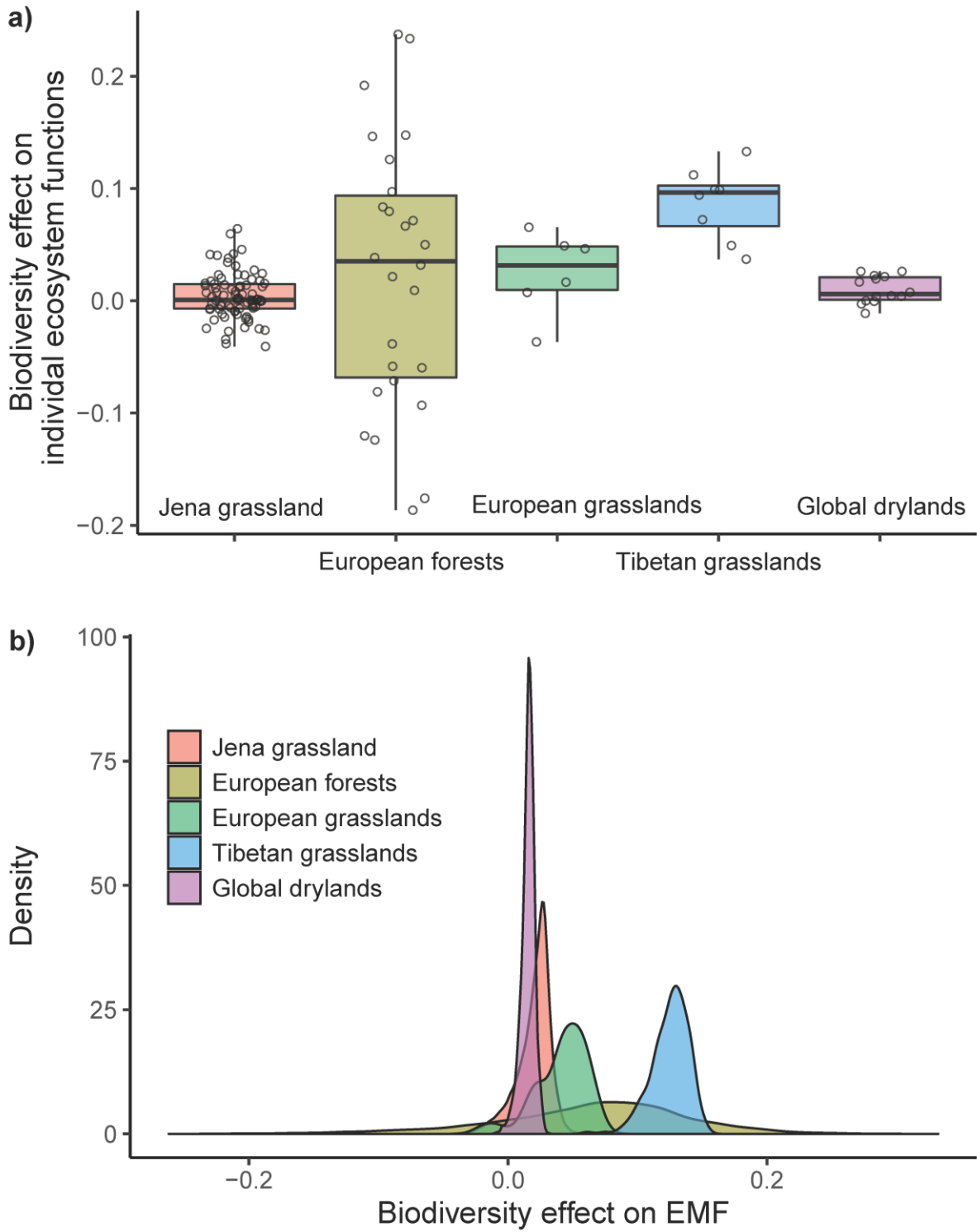
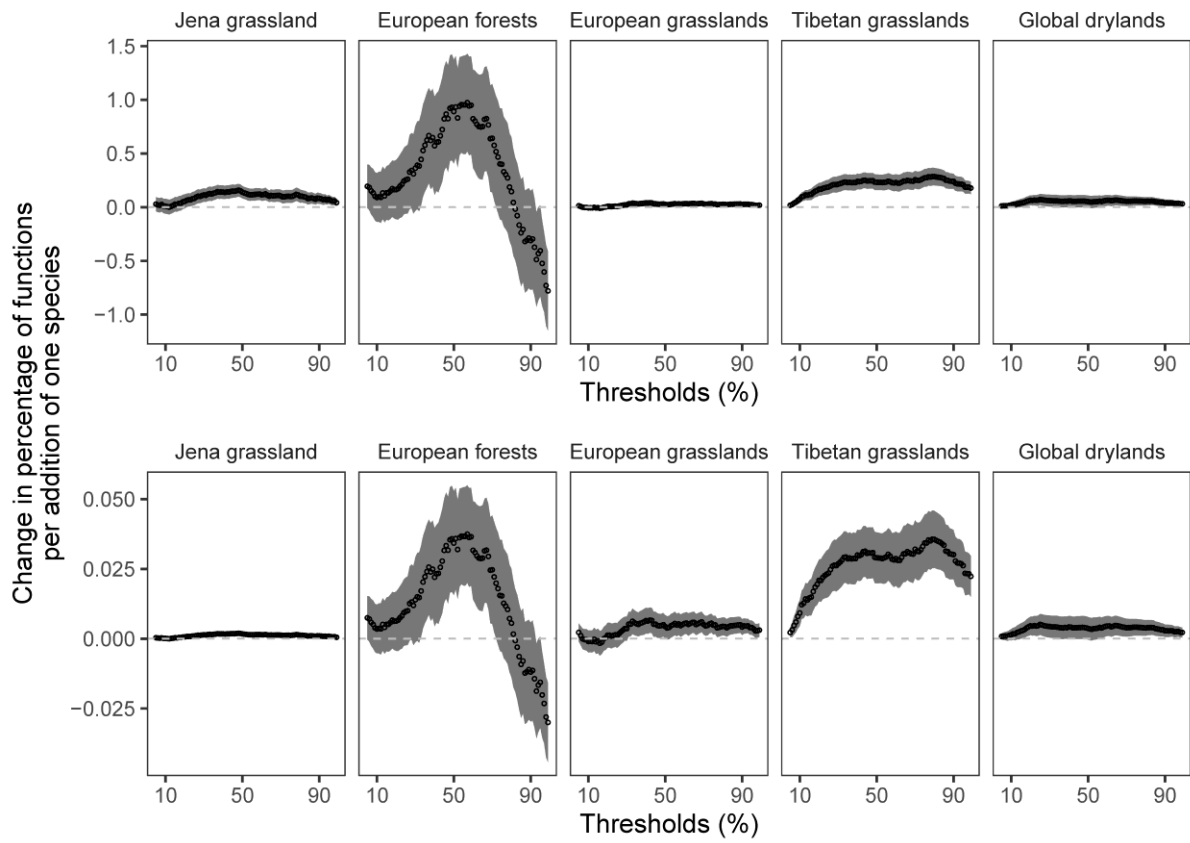


Figure 5



Accepted Manuscript