



Heriot-Watt University  
Research Gateway

## Using Species Proportions to Quantify Turnover in Biodiversity

### Citation for published version:

Yuan, Y, Buckland, ST, Harrison, PJ, Foss, S & Johnston, A 2016, 'Using Species Proportions to Quantify Turnover in Biodiversity', *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 21, no. 2, pp. 363–381. <https://doi.org/10.1007/s13253-015-0243-0>

### Digital Object Identifier (DOI):

[10.1007/s13253-015-0243-0](https://doi.org/10.1007/s13253-015-0243-0)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Journal of Agricultural, Biological, and Environmental Statistics

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Using Species Proportions to Quantify Turnover in Biodiversity

Yuan YUAN, Stephen T. BUCKLAND, Phil J. HARRISON, Sergey FOSS, and  
Alison JOHNSTON

Quantifying species turnover is an important aspect of biodiversity monitoring. Turnover measures are usually based on species presence/absence data, reflecting the rate at which species are replaced. However, measures that reflect the rate at which individuals of a species are replaced by individuals of another species are far more sensitive to change. In this paper, we propose families of turnover measures that reflect changes in species proportions. We study the properties of our measures, and use simulation to assess their success in detecting turnover. Using data on the British farmland bird community from the breeding bird survey, we evaluate our measures to quantify temporal turnover and how it varies across the British mainland.

Supplementary materials accompanying this paper appear on-line.

**Key Words:** Biodiversity; Breeding bird survey; Species abundance distribution; Turnover measures.

## 1. INTRODUCTION

Perhaps appropriately, there is a great diversity of measures for quantifying biodiversity (Pielou 1975; Krebs 1989; Magurran 2004). Choice of measure depends both on the type of data available and on the questions being asked. Maurer and McGill (2010) reviewed a large number of indices to measure species diversity, and gave a list of indices to measure different aspects of diversity, such as richness, evenness, dominance and rarity. This paper concentrates on temporal *turnover*, which we define as the change in species proportions over time, taking into account species identity. Thus, if a community is to exhibit zero turnover, then each species should represent a constant proportion of the community. We

---

Yuan Yuan (✉) and Stephen T. Buckland School of Mathematical and Statistics, University of St Andrews, St Andrews KY16 9LZ, UK (E-mail: [yy84@st-andrews.ac.uk](mailto:yy84@st-andrews.ac.uk)). Stephen T. Buckland (E-mail: [steve@st-andrews.ac.uk](mailto:steve@st-andrews.ac.uk)). Phil J. Harrison School of Mathematical and Statistics, ArtDatabanken, Box 7007, Bäcklövav 10, 75007 Uppsala, Sweden (E-mail: [phil.harrison@slu.se](mailto:phil.harrison@slu.se)). Sergey Foss School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK (E-mail: [s.foss@hw.ac.uk](mailto:s.foss@hw.ac.uk)). Alison Johnston British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU, UK (E-mail: [alison.johnston@bto.org](mailto:alison.johnston@bto.org)).

© 2016 The Author(s). This article is published with open access at [Springerlink.com](http://Springerlink.com)  
*Journal of Agricultural, Biological, and Environmental Statistics*  
DOI: [10.1007/s13253-015-0243-0](https://doi.org/10.1007/s13253-015-0243-0)

propose families of turnover measures in this paper, and link some of them to the metrics summarized by [Martín-Fernández et al. \(1998\)](#). We further study their properties in the context of measuring turnover, and provide transformation and modification for general applications. Although we focus on temporal turnover, these measures can also be used for quantifying spatial turnover between two neighbouring locations.

Spatial turnover is closely related to the concept of beta diversity. If we consider total diversity of a region (termed gamma diversity), we can partition it into alpha diversity, which measures average diversity at locations within the region, and beta diversity, which reflects spatial heterogeneity in diversity ([Lande 1996](#)). Thus, a region with high beta diversity has a diverse range of communities, perhaps reflecting a wide variety of habitats, while a region with low beta diversity has a relatively homogeneous community of species across the region. We refer to spatial turnover as the dissimilarity between pairs of locations that are neighbours, whereas beta diversity does not require this, and can be used to compare multiple assemblages. Beta diversity covers a broader range of objectives than the spatial turnover we discuss here. [Jøst et al. \(2010\)](#) reviewed a number of indices used for measuring similarities between assemblages in the context of beta diversity. Here, we concentrate on the dissimilarity of species composition. Measuring turnover, whether spatial or temporal, is about assessing the changes between two species compositions, and such change is measured by dissimilarity, which is also referred to as differentiation, divergence or distance in different scientific fields, such as probability theory, mathematical geology and cluster analysis. Given species proportions, we propose four different families of turnover measure, all of which can be used to quantify either spatial or temporal change in diversity.

There is no simple relationship between species turnover and other measures of diversity such as richness and evenness. For example, one assemblage can have complete turnover (no species in common between two time points), but its evenness might stay the same (the species proportions might remain the same, even though they relate to different species for the two time points). The relationship between richness and turnover is more complicated; both negative and positive relationships between spatial turnover and richness have been observed in studies on spatial turnover using species–area relationships ([Clarke and Lidgard 2000](#); [Stevens and Willig 2002](#); [Koleff et al. 2003](#); [Lennon et al. 2001](#); [Lyons and Willig 2002](#)). Quantifying changes in species diversity over time or space provide valuable insights into understanding biodiversity trends.

Traditionally, turnover usually refers to spatial turnover, and is usually measured from species presence–absence data ([Rodrigues et al. 2000](#); [La Sorte and Boecklen 2005](#)). However, if available, it is more informative to use species abundance distributions to measure the compositional change over time ([Magurran 2010](#)). Measures based on the species abundance distribution or on species proportions are more informative and sensitive to biodiversity changes than measures based on presence/absence data. When we use species abundance distributions to measure turnover, turnover is evaluated by quantifying the rate at which individuals of one species are being replaced by individuals of another species.

The U.K. Breeding Bird Survey (BBS, [Risely et al. 2013](#)) has been conducted annually since 1994 on a stratified random sample of 1 km squares. Sites are surveyed twice a year by volunteers walking along two parallel 1 km transect lines, and recording any adult birds detected. [Harrison et al. \(2014\)](#) fitted models to these data to allow estimation of abundance

in any 1 km square on the British mainland for any species with adequate data. We use those estimates for farmland species to estimate temporal turnover in the British farmland bird community, and study how it varies spatially. The BBS is undertaken by the British Trust for Ornithology (BTO) and jointly funded by the BTO, the Joint Nature Conservation Committee and the Royal Society for the Protection of Birds. The BBS data are available through the British Trust for Ornithology's standard data request procedure (see <http://www.bto.org/research-data-services/data-services/data-and-informationpolicy>).

## 2. BACKGROUND

Before introducing families of turnover measures, we briefly review indices that have been used for, or are related to, measuring turnover. Most have been used for measuring beta diversity, i.e. assessing compositional similarity between assemblages. Jøst et al. (2010) gave a list of distance-based similarity measures, some of which can be expressed by the  $L^p$  distance-based measures proposed in Sect. 4.1. Some of the entropy-based similarity measures proposed by Jøst et al. (2010) have forms similar to the one introduced in Sect. 4.4. All of the indices that we propose can be used to measure either spatial or temporal turnover. However, spatial and temporal turnover measures have different interpretations. As pointed out by Magurran (2010), spatial heterogeneity (beta diversity) typically corresponds to different subsets of individuals, whereas in temporal studies, we follow a single community of individuals over time.

There are essentially three different methods to study turnover (Chao et al. 2006; Jøst et al. 2010). First, incidence-based and abundance-based similarity measures are widely used, and they usually share similar forms. Most studies that use them are in the context of beta diversity. When there is only one assemblage and a time series of observations on it, these indices can also be used to measure temporal change in diversity. Incidence-based indices use presence–absence data, while abundance-based indices use counts or abundances. These counts or abundances are often scaled to give species proportions.

Second, spatial turnover can also be evaluated by studying the species–area relationship, which is the relationship between species richness (i.e. the number of species in an assemblage) and the size of the area that the assemblage occupies. Power and logarithmic functions have been used to estimate how the number of species increases as the size of the sampled area increases (Conner and McCoy 1979; Lennon et al. 2001; Rosenzweig 1995). The species–time relationship, a temporal analogue of the species–area relationship, is used to evaluate the dependence of species richness on temporal scale (Grinnell 1922; Preston 1960; Rosenzweig 1995). Power functions (Adler and Lauenroth 2003; Hadly and Maurer 2001) and logarithmic functions (Rosenzweig 1995; White 2004) have been used to estimate the species–time relationship and hence the temporal turnover of species.

Third, the species range shifts, i.e. changes in species' ranges over time, are estimated to quantify species turnover. As for incidence-based indices, presence–absence data are typically used, and turnover is measured by modelling species ranges, thus allowing the species extinctions and colonizations in the survey area to be estimated (Nenzén and Araújo 2011; Thuiller et al. 2005; Lawler et al. 2009). In the study of Island Biogeography theory,

MacArthur and Wilson (1967) defined species turnover as the number of species eliminated and replaced per unit time, and focus on equilibrium state in terms of immigration and extinction in an island population. In applications, we often take the predicted future range of each species based on modelling climate or land use changes, and the consequent impact on the species.

Both incidence-based and species–area relationship methods use presence–absence data. However, ‘absence’ typically equates to ‘not recorded’, which may be a consequence of failure to detect the species when it is present, rather than genuine absence. In addition, rare species are more difficult to detect and may be under-represented relative to more abundant species. For the method based on species range shifts, change in species’ ranges tends to be slow and difficult to detect, as species are typically sparsely distributed at the edge of their range, so that the problem of undetected presence is greater. However, species proportions change almost continuously. Methods based on species proportions are more sensitive to changes in the community than those based on changing ranges. We therefore concentrate on turnover measures that use the species proportion vector, derived from the estimated species abundance distributions over time or space. We introduce notation in Sect. 3.1 and some key transformations in Sect. 3.2. We give a list of criteria and discuss their importance in measuring turnover in Sects. 3.3 and 3.4. We then introduce four families of turnover measure (Sect. 4). We summarize the proposed turnover measures in tables (Web Appendix A) and compare them through a simulation study (Web Appendix B). In Sect. 5, we apply our measures to quantify turnover in the farmland breeding bird community of the British mainland for the period 1994 to 2011. Finally, we summarize our findings from the simulation study (Sect. 6) and discuss the choice of different turnover measures (Sect. 7).

### 3. PRELIMINARY INFORMATION

#### 3.1. NOTATION

We use  $\mathbf{p} = (p_1, \dots, p_K)$  to denote the vector of species proportions, where  $p_k \geq 0$ ,  $\sum_{k=1}^K p_k = 1$ , and  $K$  is the number of species. Note that some  $p_k$  may be zero. However, zero values are problematic for measures involving the ratio of  $p_k$ ’s, which includes several of our indices. In practice, therefore,  $K$  might be taken to be the number of species for which  $p_k > 0$ . For regional surveys, we may be interested in quantifying temporal turnover by location, so that spatial variation in temporal turnover can be assessed. In this case, even common species are likely to have observed species proportions of zero at some locations. In this circumstance, spatio-temporal models can be fitted to the data for each species, and species proportions calculated from the predicted abundances at each location (Harrison et al. 2014). Temporal turnover can then be evaluated at each location using all those species with a predicted abundance exceeding zero at that location (Harrison et al. 2015; Yuan et al. n.d.), even though some of those species may not have been recorded at that location. The spatio-temporal modelling also reduces the sampling error in estimated species proportions.

For simplicity, we assume  $p_k > 0$  for any species  $k$ , so that all families of turnover measures proposed here are well defined. Therefore,  $\mathbf{p}$  is considered in an open simplex, denoted by  $\mathbb{P}_+^{K-1} = \left\{ p_1, \dots, p_K \mid \sum_{k=1}^K p_k = 1 \text{ and } \forall k \ p_k > 0 \right\}$ . We assume that  $\mathbf{p}$  is a

known probability distribution vector, although to evaluate turnover measures in practice, we would need to replace it by an estimate  $\widehat{\mathbf{p}}$ . Turnover measures assess the difference between two probability distribution vectors, which we denote by  $\mathbf{p}_1$  and  $\mathbf{p}_2$  at two different time points. Thus  $\mathbf{p}_i$  is the species proportion vector at time point  $i$ ,  $\mathbf{p}_i = (p_{i1}, \dots, p_{iK})$ , for  $i = 1, 2$ . Measures can also be defined that are not a function of the species proportion vectors. For example, we might use the species density vector. In such cases, we use  $\mathbf{p}^*$  to denote the vector, which is no longer a probability distribution vector:  $p_k^* > 0, \forall k$  as before, but  $\sum_{k=1}^K p_i^* \in (0, \infty)$ .

### 3.2. SOME USEFUL TRANSFORMATIONS

We list three transformations on the open simplex  $\mathbb{P}_+^{K-1}$ , and all of them are one-to-one.

1. *Transform into a positive sphere* In order to use the distributional theory that has been established for the sphere (Stephens 1982; Stanley 1990), the simplex  $\mathbb{P}_+^{K-1}$  needs to be transformed to a sphere, and such a transformation is denoted by  $\mathcal{S}(\mathbf{p})$ ,

$$\mathcal{S}(\mathbf{p}) := (\sqrt{p_1}, \dots, \sqrt{p_K}). \quad (1)$$

Using this transformation,  $\mathbb{P}_+^{K-1}$  is transformed into a positive sphere, denoted by  $\mathbb{S}_+^{K-1} = \left\{ (\sqrt{p_1}, \dots, \sqrt{p_K}) \mid p_k > 0, \sum_{k=1}^K p_k = 1 \right\}$ . The transformed species proportion vector is no longer a probability distribution vector after transformation  $\mathbb{S}_+^{K-1}$ , because  $\sum_{k=1}^K \sqrt{p_k} \neq 1$ .

2. *Transform using a weight vector* We may wish to assign different levels of importance to different species. Such importance can be related to, for example, the commercial value of different species in the fishing industry, different functional roles in the ecosystem, or genetic distance of a species from others in the community. The importance of each species is represented by a weight. Let  $\mathbf{w}$  denote the weight vector, where  $\mathbf{w} \in \mathbb{P}_+^{K-1}$  with  $\sum_{k=1}^K w_k = 1$ . For any  $\mathbf{w}$ , the weight transformation,  $F_{\mathbf{w}}(\mathbf{p})$ , is a simplex operation (a transformation from  $\mathbb{P}_+^{K-1}$  to  $\mathbb{P}_+^{K-1}$ ),

$$F_{\mathbf{w}}(\mathbf{p}) := \left( \frac{p_1 w_1}{\langle \mathbf{p}, \mathbf{w} \rangle}, \dots, \frac{p_K w_K}{\langle \mathbf{p}, \mathbf{w} \rangle} \right) \quad (2)$$

where  $\langle \mathbf{p}, \mathbf{w} \rangle = \sum_{k=1}^K p_k w_k$  is the scalar product of  $\mathbf{p}$  and  $\mathbf{w}$ .

Note that when combining the above two transformations  $\mathcal{S}$  given by (1) and  $F_{\mathbf{w}}$  given by (2), we get a transformation from a simplex to a weighted positive sphere. Let  $\mathcal{S}_{\mathbf{w}}$  denote the resulting transformation from  $\mathbb{P}_+^{K-1}$  to  $\mathbb{S}_+^{K-1}$  using the weight vector  $\mathbf{w}$ :

$$\mathcal{S}_{\mathbf{w}}(\mathbf{p}) := \left( \sqrt{\frac{p_1 w_1}{\langle \mathbf{p}, \mathbf{w} \rangle}}, \dots, \sqrt{\frac{p_K w_K}{\langle \mathbf{p}, \mathbf{w} \rangle}} \right). \quad (3)$$

After transformation  $F_{\mathbf{w}}(\mathbf{p})$  as defined in (2), the species proportion vector is still a probability distribution vector. However, this is not the case for  $\mathcal{S}_{\mathbf{w}}(\mathbf{p})$  defined in (3).

3. *Centred log-ratio (clr) transformation* It can be useful to transform proportions to the real line. The log-ratio transformation,  $\log(p_k/p_K)$ ,  $k = 1, \dots, K - 1$ , is used to transform  $\mathbf{p}$  in the simplex  $\mathbb{P}_+^{K-1}$  into data in  $\mathbb{R}^{K-1}$ . However, this log-ratio transformation affects each element of  $\mathbf{p}$  differently because of its dependence on the (arbitrary) ordering of the specie in  $\mathbf{p}$ . In order to obtain a one-to-one transformation and a symmetric operation for all  $K$  components in  $\mathbf{p}$ , Aitchison (1982, page 79) proposed the clr transformation which uses the geometric mean as the divisor. Let  $g(\mathbf{p})$  denote the geometric mean of  $\mathbf{p}$ , i.e.  $g(\mathbf{p}) = \left(\prod_{k=1}^K p_k\right)^{1/K}$ . The clr transformation is defined as  $\text{clr}(\mathbf{p}) := \left(\log \frac{p_1}{g(\mathbf{p})}, \dots, \log \frac{p_K}{g(\mathbf{p})}\right)$ . The species proportions after applying  $\text{clr}(\mathbf{p})$  lie in  $\mathbb{R}$ . Therefore, the transformed vector is not a distribution vector.

### 3.3. PROPERTIES OF TURNOVER MEASURES

A turnover measure,  $d(\mathbf{p}_1, \mathbf{p}_2)$ , is considered as a function from  $\mathbb{P}_+^{K-1} \times \mathbb{P}_+^{K-1} \rightarrow \mathbb{R}$ . Such a function can be used to measure either temporal or spatial turnover. The form of the function stays the same but the arguments are different: for temporal turnover,  $\mathbf{p}_i$ ,  $i = 1, 2$ , is the species proportion vector at time  $i$  (in a given location or region), while for spatial turnover,  $\mathbf{p}_i$ ,  $i = 1, 2$ , is the species proportion vector at location  $i$  (at a given time point). Whether spatial or temporal,  $d(\mathbf{p}_1, \mathbf{p}_2)$  is a metric if it satisfies the following three properties, but the relative importance of the properties differs for the two cases.

1. *Positive definiteness*  $d(\mathbf{p}_1, \mathbf{p}_2) > 0$  for any  $\mathbf{p}_1 \neq \mathbf{p}_2$ , and  $d(\mathbf{p}_1, \mathbf{p}_2) = 0$  if and only if  $\mathbf{p}_1 = \mathbf{p}_2$ . For turnover measures, the positive definiteness means that the minimum value of a turnover measure is zero, and a turnover measure equals zero if and only if there is no change in species composition.
2. *Symmetry*  $d(\mathbf{p}_1, \mathbf{p}_2) = d(\mathbf{p}_2, \mathbf{p}_1)$ , for any  $\mathbf{p}_1 \neq \mathbf{p}_2$ . Symmetry is necessary when two assemblages or communities are compared for turnover. However, symmetry is not necessarily required for measuring temporal turnover, for which it is natural to follow the temporal order.
3. *Triangle inequality*  $d(\mathbf{p}_1, \mathbf{p}_3) \leq d(\mathbf{p}_1, \mathbf{p}_2) + d(\mathbf{p}_2, \mathbf{p}_3)$ . For both spatial and temporal turnover measures, it is preferred to have the triangle inequality. In the case of temporal turnover, for example, when comparing the turnover for three consecutive time points, the turnover between time point 1 and time point 3 should be no larger than the sum of turnover of time point 1 versus time point 2 and of time point 2 versus time point 3.

In addition to the above three properties, we note that permutation invariance is another property that is important for turnover measures. For a turnover measure, permutation invariance means that changing the sequence of the species in  $\mathbf{p}$  does not change the result. As all the measures discussed in this paper are permutation invariant, we do not list it above. It is also important to consider the above three properties for a turnover measure that is not defined on probability distribution vectors. The species proportion vector is obtained simply by scaling a density or abundance vector. For a vector  $\mathbf{p}^*$ , if we can find a constant  $c > 0$

such that  $\mathbf{p}^* = c \mathbf{p}$ , we say that  $\mathbf{p}^*$  is equivalent to  $\mathbf{p}$ . Given this definition, the species density vector,  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_K)$ , is equivalent to a species proportion vector,  $\mathbf{p}$ , because  $\mathbf{p} = \boldsymbol{\lambda} / \sum_{k=1}^K \lambda_k$ . Some measures have the above three properties not only for  $\mathbf{p}$ , but also for any of its equivalent vectors. When a measure is equivalently defined on  $\mathbf{p}$  and all its equivalent vectors  $\mathbf{p}^*$ , we use  $d(\mathbf{p}_1^*, \mathbf{p}_2^*)$  to denote the turnover measure.

### 3.4. FURTHER PROPERTIES OF TURNOVER MEASURES

It is important to examine the properties of turnover measures and understand the advantage of having these properties in applications. Jøst et al. (2010) suggested three basic properties for any ecologically useful measure of similarity and discussed these properties in the context of measuring beta diversity. Aitchison (1992) gave a list of criteria for measures to assess compositional difference in geological applications. Although concerned with a different application, their summary sheds light on the properties of turnover measures, which we discuss here. In this section, we discuss the properties of turnover measures in the context of measuring turnover as a ‘distance’ or ‘divergence’ between two probability distributions of frequencies specified by two different species proportion vectors  $\mathbf{p}_1$  and  $\mathbf{p}_2$ . In Sect. 3.3, we listed three basic properties for a turnover measure to be a metric. The following lists another two useful properties:

4. *Scale invariance*  $d(a \mathbf{p}_1, b \mathbf{p}_2) = d(\mathbf{p}_1, \mathbf{p}_2)$ , for any constants  $a > 0$  and  $b > 0$ . If a turnover measure is scale invariant, then  $\mathbf{p}_1$  and  $\mathbf{p}_2$  do not have to be probability distribution vectors, i.e.  $\sum_{k=1}^K p_{1k}$  and  $\sum_{k=1}^K p_{2k}$  do not necessarily have to be one. A scale invariant turnover measure remains unchanged as long as the species compositional vectors are proportional to the species probability vector, as is the case for species density and abundance vectors.
5. *Insensitivity to coordinate scaling*  $d(\mathbf{p}_1, \mathbf{p}_2) = d(\mathcal{C} \circ \mathbf{p}_1, \mathcal{C} \circ \mathbf{p}_2)$ , for any coordinate scaling function  $\mathcal{C}$ , which is an operation defined on vector  $\mathbf{x}$ . Let  $\mathbf{y}$  denote the vector after applying  $\mathcal{C}$  to  $\mathbf{x}$ , then  $\mathbf{y} = \mathcal{C} \circ \mathbf{x} = (\mathcal{C}_1 x_1, \dots, \mathcal{C}_K x_K)$ , and for each  $k$ ,  $y_k = \mathcal{C}_k x_k$ . Note that for a turnover measure, if the difference between each component in  $\mathbf{p}_1$  and  $\mathbf{p}_2$  is measured in the form of ratios, i.e.  $p_{1k}/p_{2k}$  for any species  $k$ , then this turnover measure is insensitive to any coordinate scaling.

Scale invariance is a special case of insensitivity to coordinate scaling, as it is the same scaling factor applied to each coordinate, i.e.  $\mathcal{C}_k$  in the coordinate scaling function is constant for any  $k$ . We consider properties 1 to 3 of Sect. 3.3 together with these two further properties in the context of the proposed turnover measures (see Web Appendix A for details).

### 3.5. THE ADVANTAGE OF BEING INSENSITIVE TO COORDINATE SCALING FOR TURNOVER MEASURES

It is rare to have perfect detection in wildlife abundance surveys. Therefore, we need to apply detection probabilities to the survey data to estimate abundance. Such a process can be considered as scaling each element in the species proportion vector  $\mathbf{p}$  by the inverse of the detection probability for the given species.

The coordinate scaling function is expressed as  $\mathbf{C}^{-1} = (C_1^{-1}, \dots, C_K^{-1})$ , where  $C_k$  is the detection probability for species  $k$  with  $0 < C_k \leq 1$ . The following proves that if a turnover measure is insensitive to coordinate scaling (which implies scale invariant), and if the detection probability for a given species is constant, then the measure is independent of the detection probability vector  $\mathbf{C}^{-1}$ . Using insensitivity to coordinate scaling and scale invariance in the following three steps, we have  $d(\mathbf{p}_1, \mathbf{p}_2) = d(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = d(\mathbf{C}^{-1} \circ \boldsymbol{\lambda}_1, \mathbf{C}^{-1} \circ \boldsymbol{\lambda}_2) = d(\mathbf{p}_1^{(c)}, \mathbf{p}_2^{(c)})$ , where  $\mathbf{p}^{(c)}$ ,  $i = 1, 2$ , is the species proportion vector after applying the capture probabilities. This means that for species  $k$ , we have  $p_k^{(c)} = \frac{\lambda_k/C_k}{\sum_{s=1}^K (\lambda_s/C_s)}$ . Therefore, provided that for any given species, the detection probability does not change over time, so that  $\mathbf{C}$  is constant, we do not need to be able to estimate detection probability for a turnover measure that is scale invariant and insensitive to coordinate scaling. For surveys with standardized field procedures, such as national breeding bird surveys or long-term bottom trawl fish surveys using standard gear, it may not be unreasonable to assume that the detection probability does not change over time.

## 4. DIFFERENT FAMILIES OF TURNOVER MEASURES

### 4.1. THE $L^q$ -DISTANCE TURNOVER MEASURE AND ITS GENERALIZATIONS

It is natural to use distance to measure the difference between two vectors in  $\mathbb{P}_+^{K-1}$ . There are different forms of distance-based similarity/dissimilarity indices, and the  $L^1$  and  $L^2$  distances are most common in biodiversity applications. Ludwig and Reynolds (1988) listed a group of indices based on Euclidean distance for assessing similarity/dissimilarity between two objects, and discussed their applications in measuring difference in abundance between different sampling locations. Foster and Bills (2004) reviewed a similar group of distance-based dissimilarity indices in measuring biodiversity of fungi. Champely and Chessel (2002) generated the Euclidean dissimilarity coefficient as a function of Euclidean distance, and combined it together with principal component analysis to compare communities. Jøst et al. (2010) reviewed different distance-based similarity measures in the context of measuring beta diversity, and these measures are related to either  $L^1$  or  $L^2$  distance.

In contrast with the frequent use of  $L^1$  and  $L^2$  distances in measuring similarity or dissimilarity between communities/assemblages, we find that the  $L^q$  distance with  $0 < q < 1$  is rarely used in the biodiversity literature. Most of the distance-based measures can be unified into a class of distance-based indices using the  $L^q$  norm, where  $q$  is a positive real number. In this section, we introduce a family of measures based on the  $L^q$  distance together with a generalization when  $q = 2$ . We use  $\|\cdot\|_q$  to denote the  $L^q$ -norm of  $\mathbf{p}$ ,  $\|\mathbf{p}\|_q = \left(\sum_{k=1}^K |p_k|^q\right)^{\frac{1}{q}}$ , and for simplicity,  $\|\mathbf{p}\| = \|\mathbf{p}\|_2$ . Given the  $L^q$ -norm, we define the  $L^q$ -distance measure,  $d_q(\mathbf{p}_1, \mathbf{p}_2)$ , as

$$d_q(\mathbf{p}_1, \mathbf{p}_2) = \frac{\|\mathbf{p}_1 - \mathbf{p}_2\|_q}{\|\mathbf{p}_1\|_q + \|\mathbf{p}_2\|_q}. \quad (4)$$

1. When  $q \geq 1$ , (4) is based on the *Minkowski* distance. The numerator itself,  $\|\mathbf{p}_1 - \mathbf{p}_2\|_q$ , is a metric when  $q \geq 1$  (Minkowski 1896). However, (4), with the normalising denominator to ensure  $d_q(\mathbf{p}_1, \mathbf{p}_2) \in [0, 1]$ , only obeys the triangular inequality (and hence is only a metric) when  $q = 1$ . When  $q = 1$ , we have

$$d_1(\mathbf{p}_1, \mathbf{p}_2) = \frac{1}{2} \sum_{k=1}^K |p_{1k} - p_{2k}| = 1 - \sum_{k=1}^K \min\{p_{1k}, p_{2k}\}. \quad (5)$$

The distance measure given by (5) is known as the total variation distance between probability measures in probability theory, and it is also known as the *Manhattan* distance, city-block distance, the  $L^1$ -norm or the taxicab distance when constructing non-Euclidean geometries (Krause 1975). When  $q = 2$ , the numerator of (4) is the Euclidean distance between  $\mathbf{p}_1$  and  $\mathbf{p}_2$ , and

$$d_2(\mathbf{p}_1, \mathbf{p}_2) = \frac{\sqrt{\sum_{k=1}^K (p_{1k} - p_{2k})^2}}{\sqrt{\sum_{k=1}^K p_{1k}^2} + \sqrt{\sum_{k=1}^K p_{2k}^2}}. \quad (6)$$

This Euclidean-distance-based index can also be applied to presence–absence data to assess turnover. La Sorte and Boecklen (2005) used the Euclidean distance between expected and observed presence–absence vectors to evaluate the level of compositional similarity for common species in avian assemblages in North America. When  $q \rightarrow +\infty$ , only the largest changes in species proportions contribute to the  $L^q$  distance turnover measure. Clearly,  $\lim_{q \rightarrow +\infty} (p_1^q + \dots + p_K^q)^{1/q} = \max_k p_k$ .

2. Clearly,  $p_k^q \rightarrow 1$  as  $q \rightarrow 0$ , so  $\lim_{q \rightarrow 0} (p_1^q + \dots + p_K^q)^{1/q} = \infty$  if  $K \geq 0$  and at least two of the  $p_k$ 's are strictly positive. Instead, we may consider  $\lim_{q \rightarrow 0} \left( \frac{p_1^q + \dots + p_K^q}{K} \right)^{1/q}$  to shed some light on the properties of the  $L^q$  distance when  $q \in (0, 1)$ . By l'Hôpital's rule,

$$\begin{aligned} \lim_{q \rightarrow 0} \left( \frac{p_1^q + \dots + p_K^q}{K} \right)^{1/q} &= \exp \left( \lim_{q \rightarrow 0} \frac{1}{q} \log \left( \frac{p_1^q + \dots + p_K^q}{K} \right) \right) \\ &= \exp \left( \lim_{q \rightarrow 0} \frac{K}{\sum_{k=1}^K p_k^q} \sum_{k=1}^K \frac{p_k^q \log p_k}{K} \right) = \prod_{k=1}^K p_k^{1/K}. \end{aligned}$$

This means that if we use  $\left( \sum_{k=1}^K |p_k|^q / K \right)^{1/q}$  as the  $L^q$  norm when  $0 < q < 1$ , then when  $q \rightarrow 0$ , for any species  $k$ , change in  $p_k$  contributes to the turnover measure through the  $K$ th root of the absolute difference in species proportions,  $|p_{1k} - p_{2k}|^{1/K}$ . After taking the  $K$ th root of the absolute change, the difference between abundant and rare species is much less severe compared with the limiting case when  $q \rightarrow +\infty$ . Royden (1968) pointed out that the  $L^q$  distance with  $0 < q < 1$  is less affected by extreme differences than the Euclidean distance and can therefore be more robust to

outliers. In the context of measuring turnover, the outliers can be thought of as species that contribute most to the turnover, i.e. the species that have the greatest change in their proportion. The absolute change in proportions will tend to be larger for abundant species than for rare species.

3. We consider a generalization of the  $L^2$ -distance measure to include similarity among species. We write (6) in a quadratic form as  $\sqrt{(\mathbf{p}_1 - \mathbf{p}_2)^T \mathbf{H} (\mathbf{p}_1 - \mathbf{p}_2)} / (\|\mathbf{p}_1\| + \|\mathbf{p}_2\|)$ , where the matrix  $\mathbf{H}$  is an identity matrix, i.e.  $h_{ks} = 1$  if  $k = s$ , and  $h_{ks} = 0$  if  $k \neq s$ . Motivated by this quadratic form of (6), we incorporate the similarity among species into (6) by using a similarity matrix  $\mathbf{Z}$  instead of  $\mathbf{H}$ ,  $d_{2S}(\mathbf{p}_1, \mathbf{p}_2) = \frac{\sqrt{(\mathbf{p}_1 - \mathbf{p}_2)^T \mathbf{Z} (\mathbf{p}_1 - \mathbf{p}_2)}}{\sqrt{\mathbf{p}_1^T \mathbf{Z} \mathbf{p}_1} + \sqrt{\mathbf{p}_2^T \mathbf{Z} \mathbf{p}_2}}$ , where the similarity matrix,  $\mathbf{Z}$ , is a  $K \times K$  matrix.  $\mathbf{Z}$  is a symmetric positive-definite matrix. The element in the  $k$ th row and  $s$ th column,  $Z_{ks}$ , quantifies the similarity between the  $k$ th and  $s$ th species, and  $Z_{ks} = Z_{sk}$  for any  $k, s \in \{1, \dots, K\}$ . We define  $0 \leq Z_{ks} \leq 1$  with  $Z_{ks} = 0$  meaning complete dissimilarity and  $Z_{ks} = 1$  meaning that two species are identical. Unlike the asymmetric similarity matrix used by [Leinster and Cobbold \(2012\)](#), we take  $\mathbf{Z}$  as symmetric positive-definite so that the product under the square root in the numerator of  $d_{2S}(\mathbf{p}_1, \mathbf{p}_2)$  is always positive as long as  $\mathbf{p}_1 \neq \mathbf{p}_2$ . When including the similarity information in the numerator, we change the denominator accordingly by incorporating  $\mathbf{Z}$  in the  $L^2$ -norm of the species proportion vector:  $\sqrt{\mathbf{p}_1^T \mathbf{Z} \mathbf{p}_1}$  is considered as an  $L^2$ -norm of  $\mathbf{p}_1$  weighted by the similarity matrix.

## 4.2. THE ANGULAR TURNOVER MEASURE

In addition to the family of distance-based measures proposed in Sect. 4.1, we can also use the angle between  $\mathbf{p}_1$  and  $\mathbf{p}_2$  to evaluate the species turnover. In this section, we propose two different groups of angular turnover measures on the basis of the space in which the angle is considered: one group is based on the angle between  $\mathbf{p}_1$  and  $\mathbf{p}_2$  in the  $\mathbb{P}_+^{K-1}$ , and the other is based on the angle in  $\mathbb{S}_+^{K-1}$  after applying the transformation (1) to  $\mathbf{p}_1$  and  $\mathbf{p}_2$ .

1. The angle in  $\mathbb{P}_+^{K-1}$  Let  $\theta$  denote the angle between  $\mathbf{p}_1$  and  $\mathbf{p}_2$  in  $\mathbb{P}_+^{K-1}$ ; see Sect. 3.1 for the definition of  $\mathbb{P}_+^{K-1}$ . The cosine of  $\theta$  is calculated as  $\cos \theta = \langle \mathbf{p}_1, \mathbf{p}_2 \rangle / (\|\mathbf{p}_1\| \|\mathbf{p}_2\|)$ . To obtain a turnover measure as a monotonically increasing function about  $\theta$ , we use  $1 - \cos \theta$  as an angular turnover measure, denoted by  $d_{\cos}(\mathbf{p}_1, \mathbf{p}_2)$ , where

$$d_{\cos}(\mathbf{p}_1, \mathbf{p}_2) = 1 - \frac{\sum_{k=1}^K p_{1k} p_{2k}}{\sqrt{\sum_{k=1}^K p_{1k}^2} \sqrt{\sum_{k=1}^K p_{2k}^2}}. \quad (7)$$

It follows that  $d_{\cos}(\mathbf{p}_1, \mathbf{p}_2) = 0$  if and only if there is no turnover, and the greater the turnover, the closer  $d_{\cos}(\mathbf{p}_1, \mathbf{p}_2)$  is to 1. Given that the angle is considered in an open simplex (i.e. for any  $k, p_k > 0$ ),  $d_{\cos}(\mathbf{p}_1, \mathbf{p}_2) < 1$ . As mentioned in Sect. 3.2, sometimes it is necessary to incorporate different weights for different species in measuring turnover, and this can be done by using the weight transformation (2). Given the

weight vector  $\mathbf{w}$ , we obtain  $d_{\cos}^{\mathbf{w}}(\mathbf{p}_1, \mathbf{p}_2) = 1 - \frac{\sum_{k=1}^K (p_{1k} p_{2k} w_k^2)}{\sqrt{\sum_{k=1}^K (p_{1k} w_k)^2} \sqrt{\sum_{k=1}^K (p_{2k} w_k)^2}}$ .

It follows that (7) is a special case of  $d_{\cos}^{\mathbf{w}}(\mathbf{p}_1, \mathbf{p}_2)$  when each species has equal weight, i.e.  $w_k = 1/K$  for all  $k$ .

2. The angle in  $\mathbb{S}_+^{K-1}$  The transformation given by (1) is to transform  $\mathbf{p}$  from the simplex  $\mathbb{P}_+^{K-1}$  to the positive sphere  $\mathbb{S}_+^{K-1}$ , so that we can use the distribution theory that has already been established for the sphere (Stephens 1982; Stanley 1990). This section introduces the idea of using the angle in  $\mathbb{S}_+^{K-1}$  because of its connection to the *Bhattacharyya* divergence measure (Bhattacharyya 1943). Although the *Bhattacharyya* divergence measure is usually referred to as a similarity measure of two probability distributions, we find that it can be derived as the angle in the positive sphere  $\mathbb{S}_+^{K-1}$  between transformed species proportion vectors.

The idea is to transform  $\mathbf{p}$  using (1), and then use the cosine of the angle in  $\mathbb{S}_+^{K-1}$  to measure turnover. Further, we incorporate weights for each species using the weight transformation given in (2). The weighting and transformation process combines (1) and (2) as  $(p_1, \dots, p_K) \mapsto \left( \sqrt{\frac{p_1 w_1}{\langle \mathbf{p}, \mathbf{w} \rangle}}, \dots, \sqrt{\frac{p_K w_K}{\langle \mathbf{p}, \mathbf{w} \rangle}} \right)$ . The cosine of the angle in  $\mathbb{S}_+^{K-1}$  is then evaluated by  $\frac{\sum_{k=1}^K \sqrt{\frac{p_{1k} w_k}{\langle \mathbf{p}_1, \mathbf{w} \rangle}} \sqrt{\frac{p_{2k} w_k}{\langle \mathbf{p}_2, \mathbf{w} \rangle}}}{\sqrt{\sum_{k=1}^K \frac{p_{1k} w_k}{\langle \mathbf{p}_1, \mathbf{w} \rangle}} \sqrt{\sum_{k=1}^K \frac{p_{2k} w_k}{\langle \mathbf{p}_2, \mathbf{w} \rangle}}}$ , where  $\langle \mathbf{p}, \mathbf{w} \rangle = \sum_{k=1}^K p_k w_k$  with a weight vector  $\mathbf{w}$ . In application,  $\mathbf{w}$  is usually defined in the simplex  $\mathbb{P}_+^{K-1}$ , i.e. for any species  $k$ ,  $w_k > 0$  and  $\sum_{k=1}^K w_k = 1$ . However, as suggested by Aitchison (1982), it is difficult to interpret  $\mathbf{w}$  on the positive sphere after being transformed from  $\mathbb{P}_+^{K-1}$ . Similar to (7), we use  $1 - \sum_{k=1}^K \sqrt{\frac{p_{1k} w_k}{\langle \mathbf{p}_1, \mathbf{w} \rangle}} \sqrt{\frac{p_{2k} w_k}{\langle \mathbf{p}_2, \mathbf{w} \rangle}}$  as a turnover measure. When using equal weights, i.e.  $w_k = 1/K, \forall k$ , we have

$$d_{\cos}^*(\mathbf{p}_1, \mathbf{p}_2) = 1 - \sum_{k=1}^K \sqrt{p_{1k} p_{2k}}. \quad (8)$$

Neither (7) nor (8) is a metric as the triangle inequality can fail. If these angular measures are considered as metrics, then they must be considered as directions on the positive sphere  $\mathbb{S}_+^{K-1}$ , in which case  $\theta$  (rather than  $1 - \cos \theta$ ) should be used for measuring turnover; the geodesic distance on a unit sphere is  $2\theta$ . Therefore,  $\theta$ , used as a turnover measure, is a metric. We instead use  $1 - \cos \theta$  because 0 then corresponds to no turnover, and 1 to complete turnover, as for the distance-based measures introduced in Sect. 4.1.

### 4.3. PAIRWISE TURNOVER MEASURES

By using pairwise comparisons, we obtain indices that are more sensitive to changes in less abundant species. We introduce two pairwise measures in this section: pairwise angular and pairwise centred log-ratio.

The pairwise angular measure uses angles between each of the  $K(K-1)/2$  pairs of species to measure the spatial or temporal changes in species composition. Let  $k$  and  $s$  denote any two species ( $k, s \in \{1, \dots, K\}; k \neq s$ ). Consider their species proportions

$(p_k, p_s)$  in an x-y plot, with the x-axis corresponding to species  $k$  and the y-axis to species  $s$ . Let  $\theta_1^{(ks)}$  and  $\theta_2^{(ks)}$  denote the angle between species  $k$  and  $s$  in time point 1 and 2, respectively. The sine and cosine of  $\theta_1^{(ks)}$  and  $\theta_2^{(ks)}$  are evaluated by  $\sin \theta_1^{(ks)} = \frac{p_{1s}}{\sqrt{p_{1k}^2 + p_{1s}^2}}$ ,  $\sin \theta_2^{(ks)} = \frac{p_{2s}}{\sqrt{p_{2k}^2 + p_{2s}^2}}$ ,  $\cos \theta_1^{(ks)} = \frac{p_{1k}}{\sqrt{p_{1k}^2 + p_{1s}^2}}$ , and  $\cos \theta_2^{(ks)} = \frac{p_{2k}}{\sqrt{p_{2k}^2 + p_{2s}^2}}$ . We then evaluate  $\sin(\theta_1^{(ks)} - \theta_2^{(ks)}) = \sin \theta_1^{(ks)} \cos \theta_2^{(ks)} - \cos \theta_1^{(ks)} \sin \theta_2^{(ks)}$ . If there is no turnover between species  $k$  and  $s$  from time point 1 to time point 2, then there is no difference between  $\theta_1^{(ks)}$  and  $\theta_2^{(ks)}$ , and it follows that  $\sin(\theta_1^{(ks)} - \theta_2^{(ks)}) = 0$ . We use the number of pairs of species,  $K(K-1)/2$ , to scale the sum of absolute values over all unique pairs, i.e.  $\frac{1}{2} \sum_{k=1}^K \sum_{s=1}^K |\sin(\theta_1^{(ks)} - \theta_2^{(ks)})|$ . As a result, the measure does not depend on  $K$ , and has range  $[0, 1]$ . The pairwise angular measure  $d_{\sin}$  is defined as

$$d_{\sin}(\mathbf{p}_1, \mathbf{p}_2) = \frac{1}{K(K-1)} \sum_{k=1}^K \sum_{s=1}^K \frac{|p_{1k}p_{2s} - p_{1s}p_{2k}|}{\sqrt{p_{1k}^2 + p_{1s}^2} \sqrt{p_{2k}^2 + p_{2s}^2}}. \quad (9)$$

Note that we can also use tangent to replace sine in (9), i.e.  $\tan(\theta_1^{(ks)} - \theta_2^{(ks)})$  instead of  $\sin(\theta_1^{(ks)} - \theta_2^{(ks)})$ ; we just need the trigonometric function to be monotonically increasing between  $-\pi/2$  to  $\pi/2$ . Whichever trigonometric function we use, it is not possible to have the insensitivity to coordinate scaling described in Sect. 3.4. However,  $d_{\sin}$  given by (9) does satisfy the first two properties of a metric. Given the inequality  $|\sin(\alpha + \beta)| \leq |\sin \alpha| + |\sin \beta|$  for any  $\alpha$  and  $\beta$ , it is easy to prove that  $d_{\sin}$  also satisfies the triangular inequality and therefore  $d_{\sin}$  is a metric.

The pairwise centred-log-ratio measure is based on the following metric introduced by Aitchison et al. (2000) in statistical analysis of compositional data in geology,

$$d_2^{clr}(\mathbf{p}_1, \mathbf{p}_2) = \left\{ \sum_{k=1}^K \left[ \log \frac{p_{1k}}{g(\mathbf{p}_1)} - \log \frac{p_{2k}}{g(\mathbf{p}_2)} \right]^2 \right\}^{1/2} \quad (10)$$

$$= \left\{ \frac{1}{K} \sum_{k < s} \left[ \log \frac{p_{1k}}{p_{1s}} - \log \frac{p_{2k}}{p_{2s}} \right]^2 \right\}^{1/2}. \quad (11)$$

$d_2^{clr}$  can be thought of as a pairwise measure based on (11), and thus shares the sensitivity to rare species of all pairwise measures. Based on (10), the pairwise centred-log-ratio measure is based on the Euclidean distance between species proportions transformed by the centred log-ratio function given in Sect. 3.2. It follows that the pairwise centred-log-ratio measure is a metric as well. We note that  $K$  in (11) can be considered as a scalar, and  $d_2^{clr}$  can also be written as  $\left\{ \frac{1}{2K} \sum_{k=1}^K \sum_{s=1}^K \left[ \log \frac{p_{1k}}{p_{2k}} - \log \frac{p_{1s}}{p_{2s}} \right]^2 \right\}^{1/2}$ . Similar to  $d_{\sin}$  given by (9), instead of  $K$ , we suggest using the number of pairs of species,  $K(K-1)/2$ , as the divisor. An alternative form of  $d_2^{clr}$  is defined as

$$d_2^{clrS}(\mathbf{p}_1, \mathbf{p}_2) = \left\{ \frac{1}{K(K-1)} \sum_{k=1}^K \sum_{s=1}^K \left[ \log \frac{p_{1k}}{p_{2k}} - \log \frac{p_{1s}}{p_{2s}} \right]^2 \right\}^{1/2}. \quad (12)$$

Using the number of pairs as divisor in both (9) and (12) can be thought of as taking the average across all pairs of species. This ensures comparability when  $K$  varies across regions, allowing us to compare temporal turnover at different locations. As the turnover measures are defined for species proportion vectors of the same dimension,  $K$  is constant in (12). Therefore, (12) still has all the properties that hold for (10).

#### 4.4. DIVERGENCE-BASED TURNOVER MEASURE

Pearson's  $\chi^2$  and the log-likelihood ratio test are commonly used for testing the goodness-of-fit of multinomial models. [Cressie and Read \(1984\)](#) considered these two tests as special cases of power divergence statistics. [Studený et al. \(2011\)](#) used the family of divergence measures to evaluate the degree of departure from the perfectly even abundance distribution, which serves as a null model. In this section, we derive a turnover measure based on one case of the power divergence statistics, the *Kullback–Leibler* (KL) divergence measure. We show that it has the same form as the J-divergence measure studied by [Martín-Fernández et al. \(1998\)](#), which is used for hierarchical classification. The link between our measures and those studied by [Martín-Fernández et al. \(1998\)](#) is given in Web Appendix A. Suppose we wish to test whether the observed species distribution is significantly different from the species distribution specified by a null hypothesis, denoted by  $\boldsymbol{\pi}_0 = (\pi_{01}, \dots, \pi_{0K})$ , with  $K$  species in total. Let  $\mathbf{p}_t = (p_{t1}, \dots, p_{tK})$  denote the proportion vector at  $t$ . The null hypothesis is  $H_0 : \mathbf{p}_t = \boldsymbol{\pi}_0$  versus the alternative,  $H_1 : \mathbf{p}_t \neq \boldsymbol{\pi}_0$ . The family of power divergence statistics to quantify the divergence of  $\mathbf{p}_t$  from  $\boldsymbol{\pi}_0$  is defined as  $I_\nu(\mathbf{p}_t; \boldsymbol{\pi}_0) = \frac{1}{\nu(\nu+1)} \sum_{k=1}^K p_{tk} \left\{ \left( \frac{p_{tk}}{\pi_{0k}} \right)^\nu - 1 \right\}$ , where  $\nu \in \mathbb{R}$ . One advantage of using the parametric family of divergence measures is to have a parameter  $\nu$  that controls the relative weighting given to common and rare species ([Studený et al. 2011](#)).  $\nu$  specifies different members of the test statistic in the family. When  $\nu \rightarrow -1$ ,  $I_\nu(\mathbf{p}_t; \boldsymbol{\pi}_0)$  is known as the KL divergence measure.

For the case of species distribution vectors, we concentrate on the divergence between two discrete distributions. Let  $KL(\mathbf{p}_t; \boldsymbol{\pi}_0)$  denote the KL divergence of  $\mathbf{p}_t$  from  $\boldsymbol{\pi}_0$ ,  $KL(\mathbf{p}_t; \boldsymbol{\pi}_0) = \lim_{\nu \rightarrow -1} I_\nu(\mathbf{p}_t; \boldsymbol{\pi}_0) = \sum_{k=1}^K \pi_{0k} \log \left( \frac{\pi_{0k}}{p_{tk}} \right)$ . We think of measuring turnover as measuring the degree of departure between two species proportion vectors,  $\mathbf{p}_1$  and  $\mathbf{p}_2$ . We use  $KL(\mathbf{p}_2; \mathbf{p}_1)$  to quantify the divergence of  $\mathbf{p}_2$  from  $\mathbf{p}_1$ ,  $KL(\mathbf{p}_2; \mathbf{p}_1) = \sum_{k=1}^K p_{1k} \log \left( \frac{p_{1k}}{p_{2k}} \right)$ , which is also known as the *Kullback–Leibler information number* when being used as a divergence measure between two multinomial probability distributions. Given the inequality  $p_{1k} \log \left( \frac{p_{1k}}{p_{2k}} \right) \geq \frac{1}{2} p_{1k} (p_{1k} - p_{2k})^2 \quad \forall k$  ([Rao 1973](#), p. 58), it follows that  $KL(\mathbf{p}_2; \mathbf{p}_1)$  is non-negative and  $p_{1k} \log \left( \frac{p_{1k}}{p_{2k}} \right) = \frac{1}{2} p_{1k} (p_{1k} - p_{2k})^2$  if and only if  $p_{1k} = p_{2k} \quad \forall k \in \{1, 2, \dots, K\}$ . In other words,  $KL(\mathbf{p}_2; \mathbf{p}_1) = 0$  if and only if the proportion of each species stays the same, i.e. no turnover. The upper bound for  $p_{1k} \log \left( \frac{p_{1k}}{p_{2k}} \right)$

in certain cases can be found in [Sayyareh \(2011\)](#). Clearly,  $KL(\mathbf{p}_2; \mathbf{p}_1)$  is not symmetric, i.e.  $KL(\mathbf{p}_2; \mathbf{p}_1) \neq KL(\mathbf{p}_1; \mathbf{p}_2)$ . Symmetry is a desirable property for a turnover measure. Therefore, we propose the following measure:

$$d_{KL}(\mathbf{p}_1, \mathbf{p}_2) = \frac{1}{2} \sum_{k=1}^K (p_{1k} - p_{2k}) \log \left( \frac{p_{1k}}{p_{2k}} \right). \quad (13)$$

$d_{KL}(\mathbf{p}_1, \mathbf{p}_2)$  does not satisfy the triangular inequality in general, and so it is not a metric. We note that the divergence-based index  $d_{KL}$  has the same form as the J-divergence measure ([Jeffreys 1946](#)) apart from the constant  $1/2$ .

## 5. APPLICATION TO THE BBS DATA

The BBS is conducted annually on a stratified random sample of 1 km squares (see Web Appendix C for data accessibility). Line transect surveys are carried out along two parallel 1 km lines, and detected birds are assigned to one of four categories: 0–25m from the line, 25–100m, > 100m, or flying over. We use a preference index to determine which species form the farmland community ([Newson et al. 2008](#); [Renwick et al. 2012](#); [Johnston et al. 2014](#)).

For each species, [Harrison et al. \(2014\)](#) analysed the abundance data using a generalized additive model with time point, easting, northing, elevation and habitat of each 1 km square as covariates. The models incorporate a space-time smoother using eastings, northings and time point, along with a separate smooth term for altitude and each habitat covariate ([Harrison et al. 2014](#)). We use abundance estimates of each species in each 1 km square within its assumed range (see [Harrison et al. \(2014\)](#)) for each time point to evaluate the measures proposed in Sects. 4.1, 4.2, 4.3 and 4.4. The results are plotted in Fig. 1. For a more direct comparison of the measures shown in Fig. 1, each of the nine estimated measures is represented by a different colour in each 100 km square in Fig. 2.

All measures indicate relatively high turnover in the west of Scotland and in the south-east of England. [Harrison et al. \(2015\)](#) analyse BBS data in greater depth, using three measures ( $L^1$ -distance,  $d_{\cos}$  and  $d_2^{clrS}$ ), and include 94 species in their analyses.

## 6. SIMULATION STUDY

We provide details of a simulation study to assess the power of each measure for detecting turnover in Web Appendix B. Our study shows that all measures but the  $L^q$ -distance measures with  $q < 1$  perform well in detecting turnover when there are large changes in the community. Over-dispersion has little effect on the conclusions. The mathematical form of the measures means that  $d_2^{clrS}$ ,  $d_{\sin}$  and  $d_{\cos}^*$  should be more sensitive to changes among the scarce species than are  $d_1$ ,  $d_2$  and  $d_{\cos}$ . Close inspection of the simulation results confirms this, although the differences are small. This greater sensitivity is likely to be offset at least to some degree by lower precision for these measures (as greater weight is given to scarce species with smaller sample sizes), and their inability to accommodate zero species proportions.

## USING SPECIES PROPORTIONS TO QUANTIFY TURNOVER IN BIODIVERSITY

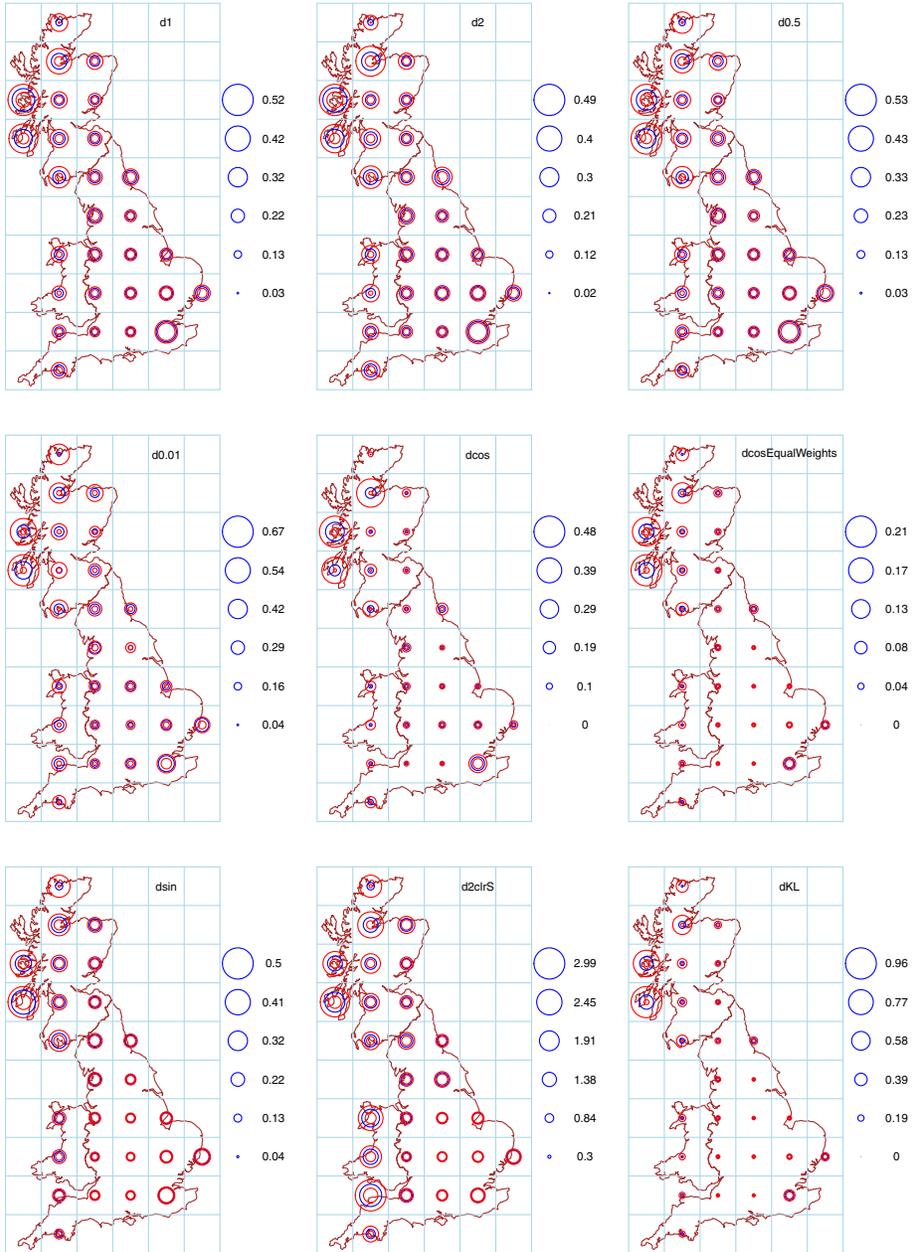


Figure 1. Turnover measures for the farmland bird community from the British breeding birds survey between 1994 and 2011. The four plots with ‘d1’, ‘d2’, ‘d0.5’ and ‘d0.01’ in the top-right corner of each plot are for the  $L^q$ -distance measures (4) with  $q = 1, 2, 0.5, 0.01$ , respectively, ‘dcos’ for  $d_{\text{cos}}$  given by (7), ‘dcosEqualWeights’ for  $d_{\text{cos}}^w$  with  $w_i = 1/K$  given by (8), ‘dsin’ for  $d_{\text{sin}}$  given by (9), ‘d2clrS’ for  $d_2^{\text{clrS}}$  given by (12) and ‘dKL’  $d_{KL}$  given by (13). Tables 1 and 2 in Web Appendix A list all of the above measures together with their properties. For each 100 km grid square, the radius of the *blue circles* indicates the size of the estimate, while 99 % confidence limits, estimated by bootstrapping (resampling the BBS survey squares within each spatial stratum), are indicated by *red circles*.

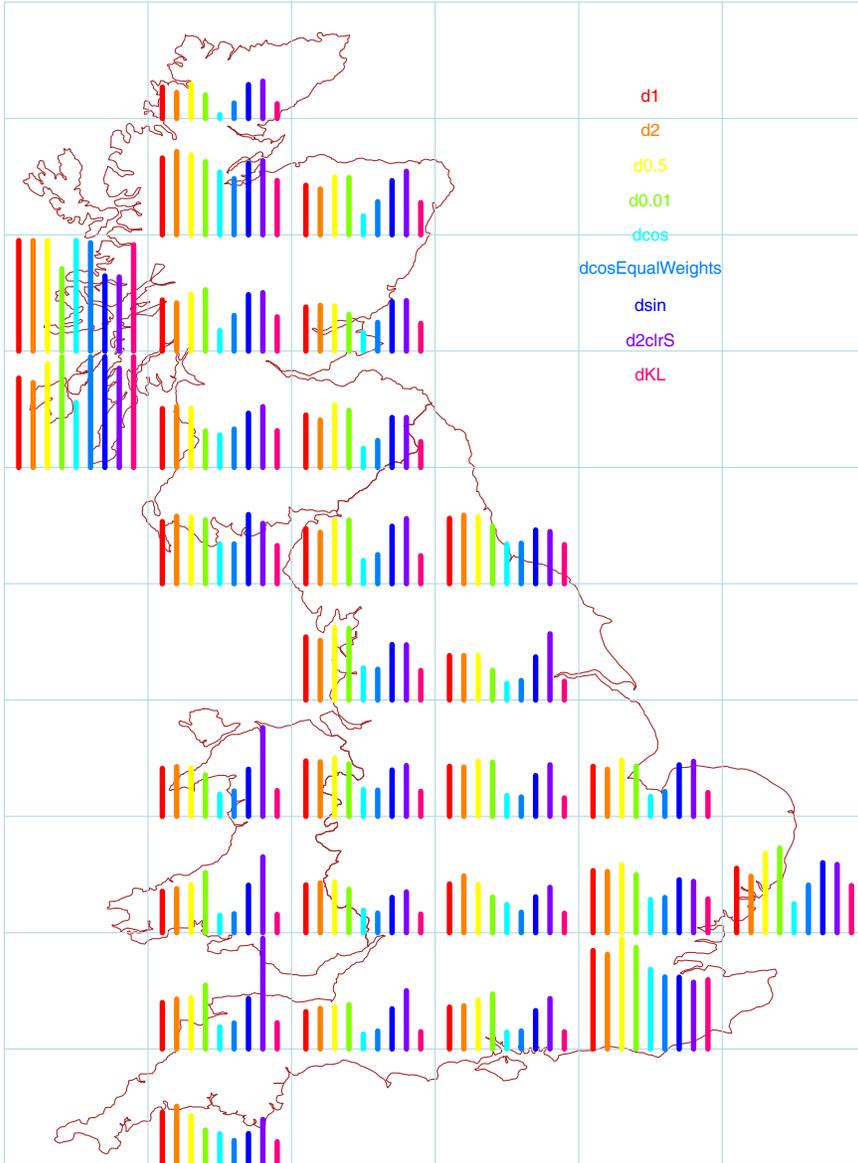


Figure 2. Plot of the estimated 9 turnover measures listed in Sect. 4 for the farmland bird community from the British breeding birds survey between 1994 and 2011. Each turnover measure is represented by a *different colour* in each 100 km square on a map of Britain. The height of the bar for each measure is divided by the maximum height for that measure across all the grid squares. Zero height corresponds to no turnover for all measures.

## 7. DISCUSSION

If a measure sensitive to changes in the scarce species is required, we recommend  $d_2^{clrS}$ , which is the only measure that satisfies all five of the desirable properties of Sects. 3.3 and 3.4. If a measure that has good precision is required, and sensitivity to changes in scarce species is not of primary interest, then we recommend  $d_1$ , which satisfies four of the

properties, and is simple, being half the sum of absolute differences in species proportions. If greater discrimination is needed between communities showing high turnover from those showing rather lower turnover, then measure  $d_2$  or  $d_{\cos}$  might be preferred. In practice, given the multivariate nature of biodiversity data, we recommend applying several turnover measures, to gain a better understanding of changes in the scarce and dominant species of the community. When the measures are applied to the BBS data, they identify high turnover in the west of Scotland and in the south-east of England. The measure  $d_2^{clrS}$  differs from the rest by also indicating high turnover in Wales.

Figure 1 shows that precision of our turnover measures is relatively poor in the west of Scotland, reflecting poor survey coverage. Nevertheless, the lower confidence limits for some of our measures in this area exceed the upper confidence limits for most of Britain, supporting our conclusion that turnover is high here. Harrison et al. (2015), in a more extensive analysis of 94 species, also reached this conclusion, and speculated that this reflects an increase in species that benefit from climate change. They attributed the high turnover in the south-east of England at least in part to a decline in scarcer specialist species.

## ACKNOWLEDGEMENTS

We are very grateful to all the volunteers who have contributed to the BBS. Yuan was funded by EPSRC/NERC grant EP/1000917/1. Harrison was funded by the Scottish Government's Centre of Expertise ClimateXChange ([www.climateexchange.org.uk](http://www.climateexchange.org.uk)).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

[Received March 2015. Accepted December 2015.]

## REFERENCES

- Adler, P. B., and Lauenroth, W. K. (2003), "The power of time: spatiotemporal scaling of species diversity," *Ecology Letters*, 6, 749–756.
- Aitchison, J. (1982), "The statistical analysis of compositional data (with discussion)," *Journal of the Royal Statistical Society: Series B*, 44, 139–177.
- Aitchison, J. (1992), "On criteria for measures of compositional difference," *Mathematical Geology*, 24(4), 365–379.
- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. A., and Pawłowsky-Glahn (2000), "Logratio analysis and compositional distance," *Mathematical Geology*, 32(3), 271.
- Bhattacharyya, A. (1943), "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.
- Champely, S., and Chessel, D. (2002), "Measuring biological diversity using Euclidean metrics," *Environmental and Ecological Statistics*, 9, 167–177.
- Chao, A., Chazdon, R. L., Colwell, R. K., and Shen, T. (2006), "Abundance-based Similarity Indices and their estimation when there are unseen species in samples," *Biometrics*, 62, 361–371.

- Clarke, A., and Lidgard, S. (2000), "Spatial patterns of diversity in the sea: bryozoan species richness in the North Atlantic," *Journal of Animal Ecology*, 69, 799–814.
- Conner, E. F., and McCoy, E. D. (1979), "Statistics and biology of the species-area relationship," *The American Naturalist*, 113, 791–833.
- Cressie, N., and Read, T. (1984), "Multinomial goodness-of-fit tests," *Journal of the Royal Statistical Society: Series B*, 46(3), 440.
- Foster, M. S., and Bills, G. F., eds (2004), *Biodiversity of fungi: Inventory and monitoring methods* Elsevier Academic Press, Burlington, USA.
- Grinnell, J. (1922), "The role of the "accidental"," *Auk*, 39, 373–380.
- Hadly, E. A., and Maurer, B. A. (2001), "Spatial and temporal patterns of species diversity in montane mammal communities of western North America," *Evolutionary Ecology Research*, 3, 477–486.
- Harrison, P. J., Buckland, S. T., Yuan, Y., Elston, D. A., Brewer, M. J., Johnston, A., and Pearce-Higgins, J. W. (2014), "Assessing trends in biodiversity over space and time using the example of British breeding birds.," *Journal of Applied Ecology*, 51, 1650–1660.
- Harrison, P. J., Yuan, Y., Buckland, S. T., Oedekoven, C. S., Elston, D. A., Brewer, M. J., Johnston, A., and Pearce-Higgins, J. W. (2015), Quantifying turnover in biodiversity of British breeding birds., in press, doi:[10.1111/1365-2664.12539](https://doi.org/10.1111/1365-2664.12539).
- Jeffreys, H. (1946), "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society A*, 184, 453–461.
- Johnston, A., Newson, S. E., Riseley, K., Musgrove, A., Massimino, D., R., B. S., and Pearce-Higgins, J. W. (2014), "Species traits explain variation in detectability of UK birds," *Bird Study*, 61, 340–350.
- Jøst, L., Chao, A., and Chazdon, R. L. (2010), Compositional similarity and beta diversity., in *Biological Diversity: Frontiers in Measurement and Assessment*, eds. A. E. Magurran, and B. J. McGill, Oxford University Press, pp. 66–84.
- Koleff, P., Gaston, K. J., and Lennon, J. J. (2003), "Measuring beta diversity for presence-absence data," *Journal of Animal Ecology*, 72, 367–382.
- Krause, E. F. (1975), *Taxicab Geometry: An Adventure in Non-Euclidean Geometry* Dover Publications, New York.
- Krebs, C. J. (1989), *Ecological Methodology* Harper and Row, Publ. Inc., New York.
- La Sorte, F. A., and Boecklen, W. J. (2005), "Temporal turnover of common species in avian assemblages in North America," *Journal of Biogeography*, 32, 1151–1160.
- Lande, R. (1996), "Statistics and partitioning of species diversity, and similarity among multiple communities," *Oikos*, 76, 5–13.
- Lawler, J. J., Shafer, S., White, D., Kareiva, P., Maurer, E. P., Blaustein, A. R., and Bartlein, P. J. (2009), "Projected climate-induced faunal change in the Western Hemisphere," *Ecology*, 3, 588–597.
- Leinster, T., and Cobbold, C. A. (2012), "Measuring diversity: the importance of species similarity," *Ecology*, 93, 477–489.
- Lennon, J. J., Koleff, P., Greenwood, J. J. D., and Gaston, K. J. (2001), "The geographical structure of British bird distributions: diversity, spatial turnover and scale," *Journal of Animal Ecology*, 70, 966–976.
- Ludwig, J. A., and Reynolds, J. F. (1988), *Statistical Ecology: a Primer of Methods and Computing* Wiley Press, New York.
- Lyons, S. K., and Willig, M. R. (2002), "Species richness, latitude, and scale-sensitivity," *Ecology*, 83, 47–58.
- MacArthur, R. H., and Wilson, E. O. (1967), *The theory of island biogeography* Princeton University Press, New Jersey.
- Magurran, A. E. (2004), *Measuring Biological Diversity* Blackwell Science, Oxford.
- Magurran, A. E. (2010), Measuring biological diversity in time (and space), in *Biological Diversity: Frontiers in Measurement and Assessment*, eds. A. E. Magurran, and B. J. McGill, Oxford University Press, pp. 85–94.
- Martín-Fernández, J. A., Barceló-Vidal, C., and Pawłowsky-Glahn, V. (1998), "Measures of difference for compositional data and hierarchical clustering methods," *The Fourth Annual Conference of the International Association for Mathematical Geology*, 2, 526–531.

## USING SPECIES PROPORTIONS TO QUANTIFY TURNOVER IN BIODIVERSITY

- Maurer, B. A., and McGill, B. J. (2010), Measurement of Species Diversity., in *Biological Diversity: Frontiers in Measurement and Assessment*, eds. A. E. Magurran, and B. J. McGill, Oxford University Press, pp. 55–65.
- Minkowski, H. (1896), “Sur Les Propriétés des nombres entiers qui sont dérivées de l’intuition de l’espace.” *Nouvelles Annales de Mathématiques, 3e série*, 15, 271–277. Also in *Gesammelte Abhandlungen*, 1. Band, XII, pp. 271–277.
- Nenzén, H. K., and Araújo, M. B. (2011), “Choice of threshold alters projections of species range shifts under climate change,” *Ecological Modelling*, 222, 3346–3354.
- Newson, S. E. and Evans, K. L., Noble, D. G., Greenwood, J. J. D., and Gaston, K. J. (2008), “Use of distance sampling to improve estimates of national population sizes for common and widespread breeding birds in the UK,” *Journal of Applied Ecology*, 45, 1330–1338.
- Pielou, E. C. (1975), “Species abundance distribution,” in *Ecological Diversity* Wiley Interscience, New York.
- Preston, F. W. (1960), “Time and space and the variation of species,” *Ecology*, 41, 612–627.
- Rao, C. R. (1973), *Linear Statistical Inference and Its Application*, 2nd edn John Wiley & Sons, New York.
- Renwick, A. R., Johnston, A., Joys, A., Newson, S. E., Noble, D. G., and Pearce-Higgins, J. W. (2012), “Composite bird indicators robust to variation in species selection and habitat specificity,” *Ecological Indicators*, 18, 200–207.
- Risely, K., Massimino, D., Newson, S. E., Eason, M. A., Musgrove, A. J., Noble, D. G., Procter, D., and Baillie, S. R. (2013), *The breeding bird survey 2012*, British Trust for Ornithology, Thetford. BTO Research Report 645.
- Rodrigues, A. S. L., Gaston, K. J., and Gregory, R. D. (2000), “Using presence absence data to establish reserve selection procedures that are robust to temporal species turnover,” *Proceedings of the Royal Society B: Biological Sciences*, 267, 897–902.
- Rosenzweig, M. L. (1995), *Species diversity in space and time* Cambridge University Press.
- Royden, H. (1968), *Real Analysis* MacMillan Publishing Co. New York.
- Sayyareh, A. (2011), “A new upper bound for Kullback-Leibler Divergence,” *Applied Mathematical Sciences*, 5(67), 3303–3317.
- Stanley, C. R. (1990), “Descriptive statistics for N-dimensional closed arrays: a spherical coordinate approach,” *Journal of Mathematical Geology*, 22, 933–956.
- Stephens, M. A. (1982), “Use of the von Mises distribution to analyze continuous proportions,” *Biometrika*, 69, 197–203.
- Stevens, R. D., and Willig, M. R. (2002), “Geographical ecology at the community level: perspectives on the diversity of new world bats,” *Ecology*, 83, 545–560.
- Studený, A. C., Buckland, S. T., Illian, J. B., Johnston, A., and Magurran, A. E. (2011), “Goodness-of-fit measures of evenness: a new tool for exploring changes in community structure,” *Ecosphere*, 2, art15.
- Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., and Prentice, I. C. (2005), “Climate change threats to plant diversity in Europe,” *Proceedings of the National Academy of Sciences*, 102, 8245–8250.
- White, E. P. (2004), “Two-phrase species-time relationships in North American land birds,” *Ecology Letters*, 7, 329–336.
- Yuan, Y., Buckland, S. T., Harrison, P. J., and Fernandes, P. G. (n.d.), “Quantifying temporal turnover in biodiversity, and how it varies spatially using the example of the International Bottom Trawl Survey data,” *in prep.* .