Ecological Indicators 45 (2014) 721–727

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Pairwise measures of species co-occurrence for choosing indicator species and quantifying overlap

Thomas M. Neeson[a,*], Yael Mandelik[b]

[a] Center for Limnology, University of Wisconsin, Madison, WI 53704, USA
[b] Department of Entomology, The Hebrew University of Jerusalem, PO Box 12, Rehovot IL-76100, Israel

ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form 7 May 2014
Accepted 3 June 2014

Keywords:
Bivariate covariance
Co-distribution
Mean pairwise index
Proportional similarity
Surrogate
Overdispersion
Community ecology

ABSTRACT

One of the most important ecological relationships between any two species is the degree of overlap in their distributions, i.e., their co-occurrence. Quantifying this relationship is a key step in the selection of indicator species and many other analyses in conservation biology and ecology. We derived a measure of the co-occurrence of two species based on the relative mutual information (RMI) of their distributions, and then compared its performance to three existing statistics: bivariate or binary covariance (BC), mean pairwise index (MPI), and proportional similarity (PS). To make this comparison, we measured co-occurrence values for all pairwise combinations of species collected from three communities (ground-dwelling beetles, moths, and vascular plants) in the Jerusalem Mountains and Judean Foothills, central Israel. We then used these co-occurrence values to address two different ecological problems: the challenge of identifying good indicator species, and the question of whether congeneric species co-occur more than species from different genera. We found that PS and RMI were the most reliable basis for choosing indicator species, but these two statistics differed in their error structures: PS had lower rates of type I errors (false positives), while RMI had lower rates of type II errors (false negatives). We also found that congeneric species co-occurred more often than species from different genera, but this pattern was statistically significant for only some of the measures of co-occurrence. In our analysis, then, the conclusion that we reached regarding the co-occurrence of congeneric species depended on which co-occurrence statistic was used. We therefore caution that available co-occurrence statistics should not be used interchangeably, because the ecological inferences drawn from a study may depend on the choice of co-occurrence statistic. In summary, we recommend PS and RMI as pairwise measures of species co-occurrence for investigating the reliability of biodiversity indicators and other applications in conservation biology and ecology.

1. Introduction

The measurement and analysis of species co-occurrence patterns is one of the most fundamental topics in ecology. In community ecology, co-occurrence patterns are often described at the community level, e.g., by summarizing the number of checkerboard units or other patterns in the community matrix (see Gotelli, 2000 for a review). In this paper, we focus on a different but related aspect of co-occurrence: the measurement of the degree of overlap in the distributions of two species (Veech, 2013, 2014). This problem entails quantifying the similarity of two rows in a standard community matrix, where each row represents a frequency distribution describing the presence or abundance of a species across sites. Though pairwise measures of species co-occurrence are less commonly used than community-level statistics, pairwise measurements are nonetheless an essential component of many analyses in community ecology (e.g., testing whether related species co-occur; Webb et al., 2002; Bell, 2005), conservation biology (e.g., the selection of indicator and surrogate species; Caro and O'Doherty, 1999; Caro, 2010), and biogeography (e.g., the identification of chorotypes or other bio-geographic units; Olivero et al., 2011).

Though several pairwise measures of co-occurrence exist, ecologists have little basis for choosing among these measures for the selection of indicator species and other related applications. Furthermore, given that the measurement of co-occurrence is akin to measuring the overlap of two frequency distributions, we hypothesized that measures of distributional similarity developed in other disciplines might also be useful for quantifying pairwise species co-occurrence. We were particularly interested in evaluating the

* Corresponding author. Tel.: +1 608 338 2235; fax: +1 608 265 2340.
E-mail address: neeson@wisc.edu (T.M. Neeson).

http://dx.doi.org/10.1016/j.ecolind.2014.06.006
1470-160X/

utility of mutual information, which is a mathematical measure of the amount of information that one statistical distribution provides about another (Manning and Schutze, 1999; Hart et al., 2000; MacKay, 2005). Given the widespread use of mutual information in other domains and its appealing mathematical properties (discussed later), we hypothesized that the mutual information of two species' distributions might be a useful measure of their co-occurrence.

In this paper, our goals were to (1) derive a measure of the co-occurrence of two species based on the relative mutual information (RMI) of their distributions, and (2) compare this statistic to three commonly used co-occurrence statistics: proportional similarity (PS; Schoener, 1970), mean pairwise index (MPI; Winston, 1995), and bivariate species covariance (BC; Bell, 2005). To make this comparison, we used these four measures of co-occurrence (RMI, BC, MPI, and PS) to address two different ecological problems: the challenge of identifying good indicator species, and the question of whether congeneric species tend to co-occur more often than species from different genera. We briefly introduce these two applications and our motivation for including them in this study.

The first application concerns the use of co-occurrence statistics in conservation biology for the selection of indicator or surrogate species. Conservation biologists use surrogate species in a variety of roles (Caro, 2010); we focus on the simplest type of indicative assessment, i.e., the use of the presence of one species as an indicator of the presence of a second, unobserved species. In this application, useful measures of co-occurrence are those that can be used to identify a species whose distribution is tightly coupled with that of another species. The majority of research in this field is empirical rather than theoretical, with highly variable results depending on the study system and scales (Hess et al., 2006; Larsen et al., 2009; Sartersdal and Gjerde, 2011). A generic measure of co-occurrence (i.e., not one derived from a specific data set) may thus afford a better basis for choosing and evaluating indicator and surrogate species.

The second application concerns the use of co-occurrence statistics to describe patterns in the co-occurrence of related species. Ecologists have long hypothesized that phylogenetic and taxonomic relationships among species can affect the structure of present-day communities (Johnson and Stinchcombe, 2007). These patterns come about because, in many cases, closely related species exhibit similar traits and habitat preferences (Johnson and Stinchcombe, 2007). If environmental filtering plays a dominant role in structuring communities, then the ecological similarity of related species might lead to a high degree of co-occurrence among these species (e.g., as in Webb, 2000; Tofts and Silvertown, 2000; Weiblen et al., 2006). On the other hand, if competition plays a more dominant role in structuring communities, then the ecological similarity of taxonomically-related species might lead to low co-occurrence among these species via competitive exclusion (e.g., as in Silvertown et al., 2001; Cavendar-Barres et al., 2006; Webb et al., 2006). Measures of co-occurrence clearly have a central role in this research area, and useful measures of co-occurrence hold the promise of providing key insights into the processes governing community assembly. Because existing measures of co-occurrence differ in their mathematical properties, we hypothesized that the choice of co-occurrence statistic might determine whether statistically significant patterns of clustering or overdispersion of related species are observed.

In this paper, we develop a measure of the co-occurrence of two species based on the relative mutual information of their distributions. To compare the usefulness of our new statistic to other, existing measures of species co-occurrence, we calculate the pairwise co-occurrence values for three different community datasets (beetles, moths, and plants, all from central Israel). We then use these co-occurrence values to investigate the two ecological questions introduced earlier: (1) Which measure of co-occurrence is the most reliable basis for choosing indicator species?, and (2) Do taxonomically-related species co-occur more often than unrelated species? Our motivation for the first analysis was the fact that indicator reliability could serve as an independent measure of how well each statistic described species co-occurrence. In the second analysis, our aim was to explore whether the ecological conclusion that we reached (i.e., whether related species co-occur) might depend on which co-occurrence statistic was used. If true, this analysis would serve as a cautionary reminder that co-occurrence statistics should not be used interchangeably in ecological analyses, and demonstrate which statistics show overlapping or dispersed results.

2. Methods

We first review existing pairwise measures of co-occurrence, and then derive a measure of co-occurrence based on the relative mutual information of two species' distributions. We then describe field survey data of beetles, moths and plants in the Jerusalem Mountains and Judean foothills, Israel, and our two research questions.

The co-occurrence of two species i and h is the similarity or overlap of the two vectors $(N_{i1},N_{i2},\ldots,N_{ir})$ and $(N_{h1},N_{h2},\ldots,N_{hr})$ in a standard community matrix, or a statistic derived from these two vectors. For presence–absence data, each element N_{ij} has value 1 if species i is present at site j, and value 0 if the species is absent. For abundance data, N_{ij} gives the number of individuals of species i at site j. Measures of co-occurrence based on the proportional distribution of each species can be calculated by first finding the total abundance of each species (i.e., the row totals) as

$$Y_i = \sum_{j=1}^{r} N_{ij}$$

and then the proportion of species i that occurs at location j as

$$p_{ij} = \frac{N_{ij}}{Y_i}$$

2.1. Existing measures of co-occurrence

We considered three existing measures of co-occurrence: proportional similarity (PS) (Schoener, 1970), bivariate species covariance (BC) (Bell, 2005), and mean pairwise index (MPI) (Winston, 1995). The proportional similarity (PS) of two species i and h over r sites is given by

$$PS_{ih} = 1 - 0.5\sum_{j=1}^{r} |p_{ij} - p_{hj}|$$

where p_{ij} is the proportion of species i at site j, and p_{hj} is the proportion of species h at site j.

The second statistic, binary or bivariate covariance (BC), is given by

$$BC\left(X_i, X_j\right) = \frac{(n_{11}n_{00} - n_{10}n_{01})}{r(r-1)}$$

where n_{11} is the number of sites (out of r total) with both species, n_{00} is the number of sites with neither species, and n_{10} and n_{01} are the number of sites with one species but not the other. Bell (2005) notes that BC is mathematically equivalent to the correlation coefficient given by Kershaw (1960) and to a modified version of the C score introduced by Stone and Roberts (1990).