Ecological Complexity xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Ecological Complexity

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

Original research article

Effect of field sampling design on variation partitioning in a dendritic stream network

Péter Sály*, Tibor Erős

MTA Centre for Ecological Research, Balaton Limnological Institute, H-8237 Tihany, Klebelsberg Kuno u. 3., Hungary

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form 21 April 2016
Accepted 9 May 2016
Available online xxx

Keywords:
Metacommunity
Fish assemblage
Species distribution modelling
Network topology
Moran's eigenvector maps (MEM)
Relative importance of space and environment

ABSTRACT

Variation partitioning is one of the most frequently used method to infer the importance of environmental (niche based) and spatial (dispersal) processes in metacommunity structuring. However, the reliability of the method in predicting the role of the major structuring forces is less known. We studied the effect of field sampling design on the result of variation partitioning of fish assemblages in a stream network. Along with four different sample sizes, a simple random sampling from a total of 115 stream segments (sampling objects) was applied in 400 iterations, and community variation of each random sample was partitioned into four fractions: pure environmentally (landscape variables) explained, pure spatially (MEM eigenvectors) explained, jointly explained by environment and space, and unexplained variance. Results were highly sensitive to sample size. Even at a given sample size, estimated variance fractions had remarkable random fluctuation, which can lead to inconsistent results on the relative importance of environmental and spatial variables on the structuring of metacommunities. Interestingly, all the four variance fractions correlated better with the number of the selected spatial variables than with any design properties. Sampling interval proved to be a fundamentally influential sampling design property because it affected the number of the selected spatial variables. Our findings suggest that the effect of sampling design on variation partitioning is related to the ability of the eigenvectors to model complex spatial patterns. Hence, properties of the sampling design should be more intensively considered in metacommunity studies.

1. Introduction

1.1. Properties of field sampling design

Properties of field sampling design set the window through which ecologists study the spatial and temporal distribution of organisms and the determining factors affecting distribution patterns. The frame of this window is the spatio-temporal scale of the study, which has three elements in ecological sampling theory. Focusing only on the spatial aspect of the scale, the grain size is the size of the sampling units (e.g., quadrates); the sampling interval is the average distance between the neighbouring sampling units; and the extent is the total area included in the investigation (Wiens, 1989; Legendre and Legendre, 2012, p. 786). Sample size, another property of sampling design, is the total number of sampling units in the sample, and it is a simple measure of the sampling effort. An additional property is the topology of the sampling units. Topology describes the geometry by which the sampling units are ecologically connected to each other. When sampling units considered being connected, researchers assume that material and individuals can move from one sampling unit to the other one (e.g., Peterson et al., 2013).

1.2. Variation partitioning

Ecologists try to reveal the mechanisms controlling the distribution of organisms by investigating their spatial distributional patterns. One of the most frequently used statistical methods for quantifying different sources of variation of communities is variation partitioning (or variance partitioning), which was introduced into the ecological methodology by Borcard et al. (1992). In a classical approach, this method uses a sites-by-species community matrix as response data, and a sites-by-environmental variables matrix and a sites-by-spatial variables matrix as explanatory data to decompose additively the total variation of the response data into four variance fractions/proportions by fitting canonical ordination models (canonical correspondence

* Corresponding author.
E-mail addresses: saly.peter@okologia.mta.hu (P. Sály), eros.tibor@okologia.mta.hu (T. Erős).

http://dx.doi.org/10.1016/j.ecocom.2016.05.002
1476-945X/

analysis [CCA] or redundancy analysis [RDA]) on the data. One of the variance fractions is the variation explained exclusively by the studied environmental variables, denoted by [*a*] in the original paper of Borcard et al. (1992). This fraction is usually considered to reflect the importance of environmental effects which could not be associated to spatial co-variation. Another variance fraction ([*c*]) is explained purely by the spatial variables, and gives estimation on community variation that has no relationship with the environmental variables included into the environmental data matrix. However, depending on the elaboration of the study, there is a possibility that this fraction incorporates some variation that would be explainable by a latent, unmeasured environmental variable. A third variance fraction ([*b*]) is explained jointly by the studied environmental and spatial variables. In this case the effects of environmental and spatial factors on community structure cannot be disentangled. The last fourth variance fraction is the unexplained residual variation [*d*].

Peres-Neto et al. (2006) improved variation partitioning by introducing the adjusted redundancy statistic or adjusted coefficient of multiple determination (R^2_{adj}). The adjusted redundancy statistic expresses the unbiased form of the variance fractions/proportions which is controlled for the number of explanatory variables in the model and the sample size.

Since its introduction, variation partitioning has become a fundamental method to infer the measure and importance of environment- and space-related mechanisms structuring communities, especially in the field of metacommunity researches. Results mirror that this measure and importance tend to vary according to the studied group of organism (e.g., Cottenie, 2005; Beisner et al., 2006; Marzin et al., 2013), ecological data type (e.g., Cushman and McGarigal, 2004; Hoeinghaus et al., 2007; Sály et al., 2011), ecosystem type (e.g., Cottenie, 2005; Heino et al., 2015; Soininen and Weckström, 2009), spatial scale of the study (e.g., Cushman and McGarigal, 2004; Declerck et al., 2011; Heino et al., 2015; Mykrä et al., 2007), study region (e.g., Cottenie, 2005) and study years (e.g., Mesquita et al., 2006).

1.3. Relationship of sampling design and variation partitioning

Differences in the study design are among the most important factors that could lead to apparently inconsistent results of variation partitioning studies. In fact, Dray et al. (2012, p. 262–263) explicitly warned that sampling design introduces an artificial spatial structure into the data in any field study. Despite this casual relevancy, only a little interest has been taken in studying systematically how sample design influences the detected spatial variation of assemblages, although many papers have highlighted the importance of certain spatial scale elements in describing the spatial structure of beta diversity (e.g., Barton et al., 2013; Heino et al., 2015; Mykrä et al., 2007; Soininen, 2015).

In two simulation studies, Smith and Lundholm (2010) and Gilbert and Bennett (2010) found that spatial configuration and sampling strategies affect the results of variation partitioning. Further, they also found that variation partitioning did not model the simulated spatial structures of the data correctly. Migration rates (i.e., dispersal), as a spatial pattern-generating mechanism, influenced both the environment- and space-related variation (Smith and Lundholm, 2010); and significant spatially explained variations were found even when the simulated data did not contain spatial component (Gilbert and Bennett, 2010).

Spatial extent, sample size and the topology of the sampling units could obviously affect the environmental and spatial variables that researches consider relevant to describe the spatial variation of assemblages. In many researches, these explanatory variables are identified via a forward selection procedure (Blanchet et al., 2008) prior to variation partitioning. Although, the adjusted form of the variation proportions (Peres-Neto et al., 2006) takes the number of the explanatory variables into account which helps to compare the results of different studies, the effect of the sampling design properties on the number of the relevant (i.e., selected) explanatory variables has not been examined yet.

For stream-dwelling organisms like fish and aquatic molluscs that have no capacity for terrestrial movement, dispersal connectivity among habitats is completely determined by the physical dendritic structure of the stream network (Fagan et al., 2009), hence topology, beside the dispersal ability of the animals, can be supposed to play a prominent role in their spatial dynamics. The importance of topology of dendritic stream networks has been studied in connection with, for example, fish dispersal (Hitt and Angermeier, 2008, 2011) and in the context of the distance-decay similarity relationship for aquatic invertebrates (e.g., Brown and Swan, 2010; Cañedo-Argüelles et al., 2015), but the relationship between the topology of the effectively sampled locations of a dendritic network and the space-related community variation is still little known. In fact, the behaviour of variation partitioning as a response of changes in sampling design is still uncovered; therefore we do not know which sampling design properties and variance fractions may be statistically associated to each other.

In spite of the warning results mentioned above and the lack of a solid understanding of the relationship between sampling design properties and variation partitioning, the latter has been frequently used to study the metacommunity organizations of a wide variety of taxa (e.g., Alahuhta and Heino, 2013; Baldissera et al., 2012; Buschke et al., 2015; Campbell et al., 2015; Erős et al., 2012; Fernandes et al., 2014; Göthe et al., 2013; Grönroos et al., 2013).

1.4. Aims

In this paper, we present how sampling design can affect the result of variation partitioning, and how properties of sampling design can influence the number of the selected explanatory variables and the change of the individual variance fractions in a dendritic stream network using presence-absence data of fish species. Applying simple random sampling, we focused on the specific questions as follows. (1) How does sample size (sampling effort) impact the expected value of the estimated variance fractions? Assuming a fix sample size, (2) how does the change of sample configuration influence the relative importance (i.e., rank order) of the estimated variance fractions? (3) Does the change in the sample similarity cause a proportional change in the result of variation partitioning? (4) In what extent can the change of properties of sampling design other than sample size (spatial extent, sampling interval, and topology) explain the change of the individual variance fractions and the number of explanatory variables used for partitioning? Finally, (5) How strong is the association between the amount of the unique variance fractions and the number of the selected explanatory variables used for partitioning?

2. Methods

Analyses of this study progressed through three main phases. First, environmental data were gathered and fish data were predicted by a statistical model using field survey data. Second, variation partitioning of fish data was done iteratively using simple random sampling with different sample size. Last, results of the variation partitioning were analysed statistically.

2.1. Studied stream system, environmental variables, and fish data

The studied stream system is located in Hungary (Fig. 1), and contains two small rivers, the Zagyva (179 rkm) and the Tarna (105 rkm), and their tributaries (hereafter ZT system). The