

## Variation partitioning involving orthogonal spatial eigenfunction submodels

PIERRE LEGENDRE,<sup>1,3</sup> DANIEL BORCARD,<sup>1</sup> AND DAVID W. ROBERTS<sup>2</sup>

<sup>1</sup>Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7 Canada

<sup>2</sup>Department of Ecology, Montana State University, Bozeman, Montana 59717-3460 USA

**Abstract.** When partitioning the variation of univariate or multivariate ecological data with respect to several submodels of spatial eigenfunctions (e.g., Moran's eigenvector maps, MEM) acting as explanatory data, a problem occurs: although the submodels are constructed to be orthogonal to one another, the partitioning based on adjusted  $R^2$  statistics produces nonzero values in the intersections between spatial submodels. This phenomenon is described and two solutions are proposed. The first solution is to apportion the intersection fractions proportionally to the variation explained by each submodel. The second solution consists in creating a hierarchy among the spatial submodels, in accordance with hierarchy theory. These solutions lead to new partitioning equations that are described in the Appendix. R functions are provided to carry out partitioning with respect to environmental variables and spatial eigenfunction submodels. This development is important for the correct interpretation of spatial modeling results implying explanatory environmental data as well as submodels of spatial eigenfunctions involving two or more spatial scales.

**Key words:** adjusted  $R^2$ ; asymmetric eigenvector maps (AEM); hierarchical partitioning; Moran eigenvector maps (MEM); proportional apportioning; spatial analysis; spatial eigenfunctions; variation partitioning.

### INTRODUCTION

A general method for partitioning the variation of a response data matrix with respect to two or more matrices of explanatory variables in linear models was proposed by Borcard et al. (1992). That method is well-known among ecologists: variation partitioning Venn diagrams, derived from set theory, and drawn using different styles (overlapping bars, rectangles, circles, squares, and so on), are now found in many papers in community ecology and other fields.

In early papers, the partitioning was carried out by adding and subtracting  $R^2$  statistics computed from models each involving a single explanatory matrix. For two matrices of explanatory variables for example, the fractions of variation were computed as follows:

1) First, compute the  $R^2$  statistic associated with three linear models of a single explanatory matrix, computed using multiple regression if there is a single response variable, or canonical redundancy analysis (RDA) for a multivariate table of response variables containing, for example, community composition data:

$R^2(\mathbf{Y} | \mathbf{X}_1)$  is the  $R^2$  of  $\mathbf{Y}$   
 analyzed with respect to explanatory matrix  $\mathbf{X}_1$  (1a)

$R^2(\mathbf{Y} | \mathbf{X}_2)$  is the  $R^2$  of  $\mathbf{Y}$   
 analyzed with respect to explanatory matrix  $\mathbf{X}_2$  (1b)

$R^2(\mathbf{Y} | [\mathbf{X}_1, \mathbf{X}_2])$  is the  $R^2$  of  $\mathbf{Y}$   
 analyzed by  $\mathbf{X}_1$  and  $\mathbf{X}_2$  bound together in a single matrix. (1c)

2) Then, compute the fractions of variation:

[a] =  $R^2(\mathbf{Y} | [\mathbf{X}_1, \mathbf{X}_2]) - R^2(\mathbf{Y} | \mathbf{X}_2)$  (2a)

[b] =  $R^2(\mathbf{Y} | \mathbf{X}_1) + R^2(\mathbf{Y} | \mathbf{X}_2) - R^2(\mathbf{Y} | [\mathbf{X}_1, \mathbf{X}_2])$  (2b)

[c] =  $R^2(\mathbf{Y} | [\mathbf{X}_1, \mathbf{X}_2]) - R^2(\mathbf{Y} | \mathbf{X}_1)$  (2c)

[d] = residual variation =  $1 - R^2(\mathbf{Y} | [\mathbf{X}_1, \mathbf{X}_2])$ . (2d)

Manuscript received 13 November 2011; accepted 28 November 2011. Corresponding Editor: H. H. Wagner.

<sup>3</sup> E-mail: Pierre.Legendre@umontreal.ca

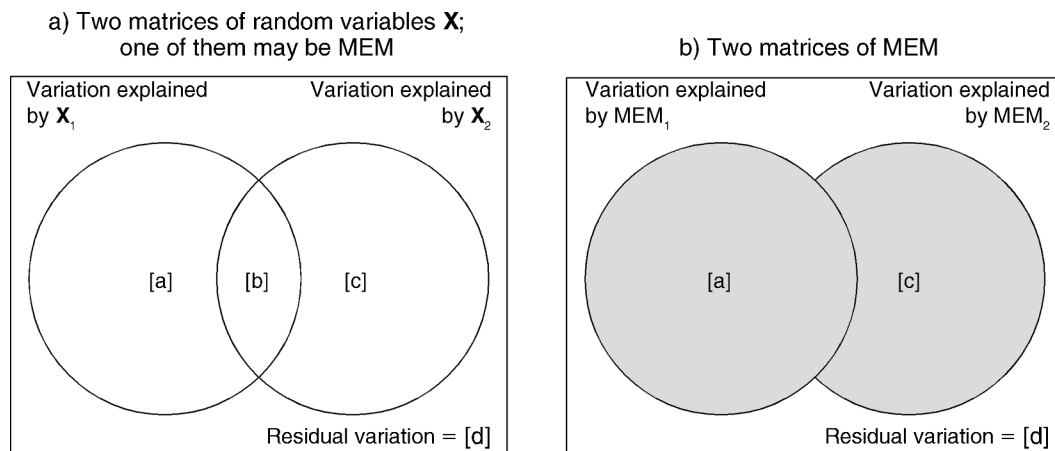


FIG. 1. Venn diagrams of variation partitioning results for (a) two matrices of random variables  $\mathbf{X}_1$  and  $\mathbf{X}_2$  (e.g., environmental variables, or a matrix of random variables  $\mathbf{X}$  and one of MEM spatial eigenfunctions); identifiers of the fractions [a] to [d] are shown; (b) two matrices of spatial MEM submodels between which a hierarchy was created. The fractions of variation explained by MEM matrices are in gray.

The notation  $(\mathbf{Y}|\mathbf{X})$  refers to the analysis (regression or RDA) of  $\mathbf{Y}$  by  $\mathbf{X}$  and  $R^2(\mathbf{Y}|\mathbf{X})$  is the  $R^2$  of that analysis. Fraction identifiers [a] to [d] are shown in Fig. 1a. That method of calculation can be extended to more than two explanatory matrices, as shown in the Appendix.

In 2006, Peres-Neto et al. showed that the  $R^2$  statistic computed in RDA is a biased estimator of the explained variation, producing increasingly high values as explanatory variables are added to the analysis, even though these variables are not linearly related to the response data under study. This phenomenon had already been shown for the  $R^2$  statistic computed in multiple regression (Ohtani 2000). Peres-Neto et al. also showed that unbiased estimates of the fractions computed in variation partitioning require that the partitioning equations be based on adjusted  $R^2$  statistics ( $R_a^2$ ; Ezekiel 1930) instead of simple  $R^2$ . To obtain  $R_a^2$  for an explanatory matrix containing random variables, i.e., variables whose values result from measurements on some types of random processes, the  $R^2$  must be corrected using its degrees of freedom, which depend in turn on the number of objects ( $n$ ) and the number of variables ( $m$ ) contained in the explanatory matrix. The adjustment is based on the fact that  $R_a^2$  must be near 0 for explanatory variables that are random (i.e., no effect) with respect to the response data  $\mathbf{Y}$ . As a result of the Peres-Neto et al. (2006) paper, the variation partitioning equations (Eqs. 2a–2d) are now based on the calculation of  $R_a^2$  statistics. In the R language, the method is implemented in function `varpart()` of the `vegan` package (*available online*).<sup>4</sup>

Spatial eigenfunction analysis is a method of growing interest for multiscale spatial analysis of ecological data.

Spatial eigenfunctions produced by these methods are called by the general name Moran's eigenvector maps (MEM; Dray et al. 2006); distance-based MEM (dbMEM), which were formerly called PCNM eigenfunctions (Borcard and Legendre 2002), belong to that group. Asymmetric eigenvector maps (AEM; Blanchet et al. 2008a, 2009) are also spatial eigenfunctions belonging to the same family. This family of methods was developed during the past 10 years. Spatial eigenfunctions can be computed for points located at regular or irregular intervals along transects or on maps. They have the properties of modeling the spatial variation at all scales encompassed in the sampling design and of being orthogonal to one another. The construction and practical use of spatial eigenfunctions is described in Chapter 7 of Borcard et al. (2011) and in Chapter 14 of Legendre and Legendre (2012).

In many studies, the eigenfunctions are divided into a number of submodels corresponding to different spatial scales of interest in a study. The eigenfunction submodels can be used as explanatory data in linear modeling, including variation partitioning. These submodels are orthogonal to one another, meaning that the scalar product of two submodels produces a null matrix. Therefore, one expects to obtain linearly independent linear models from the eigenfunction submodels, the submodels having no shared explained variation. It turns out, however, that the presently available equations for variation partitioning, based on  $R_a^2$ , create small non-null intersection values between submodels constructed from linearly independent groups of spatial eigenfunctions. These small intersection values are most often negative. That problem is revisited in the present paper.

<sup>4</sup> <http://cran.r-project.org/package=vegan>

## METHOD

How does a nonzero shared fraction of variation (e.g., fraction [b] in Fig. 1a) originate? Fraction [b] is not estimated from a linear model of the data but computed by addition and subtraction of previously computed fractions, as shown in Eq. 2b. When this equation is computed using  $R^2$  statistics, the value of [b] is 0 for two perfectly orthogonal explanatory matrices, e.g., two matrices of spatial eigenfunctions corresponding to models at different scales; values of [b] different from 0 indicate correlations between the two matrices, e.g., two matrices of environmental variables. With adjusted  $R^2$ , however, addition and subtraction of  $R_a^2$  values in Eq. 2b does not produce 0 for two matrices of orthogonal variables, but small positive or negative values.

How often are these fractions negative and how small is *small*? For example, after 10 000 repeated and independent simulations, the variation of spatially structured response variables simulated along a regular transect with 30 points, partitioned between two MEM submodels (broad-scale, fine-scale), produced 9815 negative fractions [b] (identified in Figs. 1a and 2a) out of 10 000, hence only 185 positive. The values of [b] computed from  $R_a^2$  were in the range  $[-0.0997, 0.0227]$ . In another experiment, a spatially structured response variable simulated on a regular grid with  $8 \times 12 = 96$  points was partitioned among three MEM submodels (broad-scale, middle-scale, fine-scale); the simulations produced 9959 negative fractions [d], 9651 negative fractions [e], and 9880 negative fractions [f] in 10 000 simulations (see Fig. 2a for the fraction identifiers). Fractions [d] were in the range  $[-0.0236, 0.0012]$ , fractions [e] were in the range  $[-0.0169, 0.0014]$ , and fractions [f] were in the range  $[-0.0235, 0.0012]$ .

The problem is actually not limited to matrices of spatial eigenfunctions.  $R_a^2$  statistics are not additive, so the sum of the  $R_a^2$  statistics of models computed from explanatory matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  is never strictly equal to the  $R_a^2$  of the model computed from the union of matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  even when the variables in  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are orthogonal among the groups. This problem is, however, especially annoying in studies where the variation of response data is partitioned among eigenfunction submodels corresponding to different spatial scales because these submodels are orthogonal to one another by construction.

How can we remediate this problem? In other words, using adjusted  $R^2$ , how can we obtain a partitioning of the variation of the response matrix by a combination of matrices of spatial eigenfunctions and environmental variables, without producing fractions of shared variation different from 0 between the matrices representing spatial eigenfunction submodels?

A first idea that comes to mind would be to refrain from using  $R_a^2$  statistics in analyses involving explanato-

ry matrices containing spatial eigenfunctions. However, the logic of the correction for number of variables prior to the calculation of the fractions of variation applies to spatial eigenfunctions as well: the  $R^2$  statistic increases with the number of variables in each matrix for spatial eigenfunctions, just as they do for matrices of random variables (e.g., environmental), independently of their capacity to explain the variation of the response data, so that the resulting  $R^2$  statistic overestimates the amount of explained variation.

*Proportional apportioning of the shared fractions*

A first solution is to apportion the intersection fractions proportionally to the variation explained by each submodel. The total variation explained by the combined submodels is entirely accounted for in the partition. This solution is interesting in, and applicable to, studies where no hierarchy among the spatial processes is recognized. As an example, a vegetation study where broad-scale spatial structures are considered as proxies for climatic constraints, and fine-scale patterns represent edaphic conditions, would be well served by this approach since the two classes of constraints are not conceptually nested. Calculation details for proportional apportioning are shown in the Appendix. The results are illustrated in Figs. A1–A3.

*Hierarchical partitioning of the shared fractions*

A second solution is to create a hierarchy among the orthogonal eigenfunction submodels in the analysis according to the scales they represent. In the two-submodel case, for example, we can state that the broad-scale (BS) submodel has priority over the fine-scale (FS) submodel. The consequence of that decision on the calculations is that the BS submodel is served first in the variance resource and secures for itself the small fractions found in the intersection with the FS submodel. As a result, the intersection in explained variation between the two submodels, which is of interest in partitioning involving two matrices of random variables or one matrix of random variables and a matrix of spatial eigenfunctions (Fig. 1a), disappears in Fig. 1b where two matrices of orthogonal spatial eigenfunctions (MEM) are used as explanatory variables.

The logic behind the second approach is the following: in most analyses of the spatial variation of communities, we expect environmental control processes to be at work, and we know from hierarchy theory (Simon 1962, Allen and Starr 1982, O'Neill et al. 1991) that broad-scale environmental processes, like orogenic and geomorphological processes on land or currents and winds in fluid environments, often generate broad-scale signatures in the response data. What is more difficult is to identify the signatures of smaller-scale processes like the action of biotic interactions and other forms of

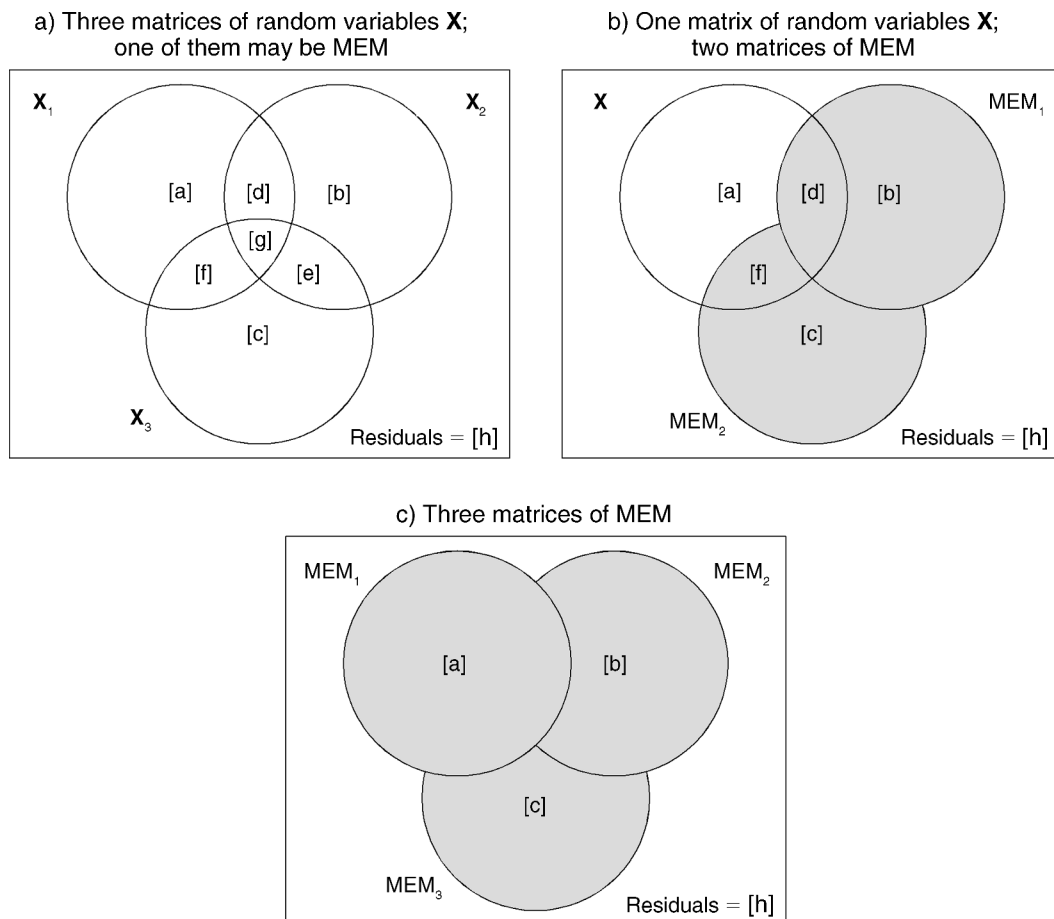


FIG. 2. Venn diagrams of variation partitioning results for three explanatory matrices: (a) classical partitioning, (b, c) results of hierarchical partitioning.

neutral processes. One must have a conservative attitude in assessing small-scale signatures of processes in order to only focus on those that are likely to be really present in data. A conservative strategy is to allow broader-scale MEM spatial models to explain variation without allowing smaller-scale submodels to interfere. The hierarchy is established on a priori grounds. The previously shared fractions become irrelevant, as are the signs that these shared fractions would have had with specific data sets. If our ecological hypothesis makes us think that we are in this kind of situation, this authorizes us to create such a hierarchy among the spatial MEM submodel matrices prior to variation partitioning. The advantage of this solution is that the  $R_a^2$  of the dominant MEM submodel is perfectly represented in the partition result, as shown in the example. If the researcher does not have a hierarchical hypothesis, he/she may use the proportional apportioning method described two paragraphs above, or simply use the traditional partitioning results that have small shared, but un-interpretable fractions.

In a study involving three explanatory matrices, one expects to find intersection fractions, which are of interest, when using either three matrices of environmental variables, or two matrices of environmental variables and a matrix of spatial eigenfunctions (Fig. 2a). If there are three matrices of spatial eigenfunctions, one can apply the hierarchy described in the previous paragraphs between eigenfunction models representing decreasing spatial scales: the broad-scale (BS) submodel may be given priority over the middle-scale (MS) submodel, which in turn has priority over the fine-scale (FS) submodel. The BS submodel is served first in the variance resource and secures for itself the small fractions found in the intersection with the MS and FS submodels. Then the MS submodel is served before the FS submodel (Fig. 2c). When there is a matrix of environmental variables and two matrices of spatial eigenfunctions, a hierarchy is created only between the two matrices representing spatial submodels (Fig. 2b): the environmental variables have shared variation with

both spatial submodels, but then the BS submodel has priority over the fine-scale (FS) submodel.

The results are fairly simple to compute. For the two explanatory matrix case, consider the fractions computed by Eq. 2, now used with  $R_a^2$  instead of  $R^2$ : for the two spatial eigenfunction case (Fig. 1b), add the value computed for [b] in Fig. 1a with the value of [a] and use the sum as the new [a] value in the hierarchy shown in Fig. 1b. For the three explanatory matrix case, Fig. 2b is obtained by adding fractions [d] and [g] of Fig. 2a to obtain the new value of [d] in Fig. 2b, and adding [b] and [e] of Fig. 2a to obtain the new value for [b] in Fig. 2b. For the three spatial submodel case shown in Fig. 2c, [a] is obtained by adding [a], [d], [f] and [g] from Fig. 2a; [b] is obtained by adding [b] and [e] from Fig. 2a; and [c] has the same value as in Fig. 2a.

The same calculations can be extended to the four explanatory matrix situation. The Venn diagrams are harder to draw (Appendix: Fig. A4), but the logic is the same. Note that in all cases, the residuals remain unchanged.

The calculation details for hierarchical partitioning involving two, three, and four explanatory matrices are given in the Appendix. The results are illustrated in Figs. 1, 2, and A4.

### Software

Functions `varpart2.MEM()`, `varpart3.MEM()`, and `varpart4.MEM()` in R can be used to carry out the calculations. They are provided in a Supplement. These functions offer the hierarchical or the proportional methods as options. In the output tables of these functions, the variation corresponding to the fractions that have disappeared is left blank although the proper notation would be the empty set ( $\emptyset$ ). These functions rely upon function `varpart()` of the *vegan* package (see footnote 4). Functions `print.varpart()` and `plot.varpart()` of the *vegan* package can be applied to the output of these functions to print partition tables and produce Venn diagrams.

Tests of significance of the fractions of variation explained by the eigenfunction submodels can be carried out as usual with function `anova.cca()` of *vegan*, using all other explanatory matrices in the analysis as covariables. The choice of the partitioning method (hierarchical or proportional) for spatial eigenfunction submodels does not matter and has no influence on the results of the tests.

### EXAMPLE ON REAL DATA

We exemplify the new procedures by applying them to the well-known oribatid mite data set that has been used in many papers dealing with variation partitioning and spatial analysis (e.g., Borcard et al. 1992, 2004, 2011, Dray et al. 2006). The data are available in the *vegan* and *ade4* R language packages as well as with the

Borcard et al. (2011) book. The mite community composition data (70 sites and 35 species) were Hellinger-transformed prior to variation partitioning (Legendre and Gallagher 2001). The three explanatory matrices involved are (1) an environmental matrix containing two quantitative variables (water content in percent and substrate density in grams per liter of dry uncompressed matter) and two matrices of dbMEM eigenfunctions that were forward-selected using the Blanchet et al. (2008b) double stopping criterion. We first selected the hierarchical solution for this example because the hypothesized processes correspond to the general framework exposed above: most environmental constraints are of the same kind, vary at broad scale, and are expected to impose coarse patterns on the community structure, while more local substrate variation and biotic interactions are expected to induce fine-scale structure within the main patterns. The broad-scale MEM submodel contains MEM 1 to 4 whereas the fine-scale submodel comprises MEM 6 to 11 and 16.

Fig. 3a shows the result of a classical partitioning and Fig. 3b the result of the hierarchical partitioning procedure. The comparison shows how giving priority to the broad-scale MEM submodel makes the interpretation easier. Note that since fractions [e] and [g] of the classical partitioning were negative (a common occurrence), the final amount of variation explained by the broad-scale MEM submodel, as well as the fraction common to the environmental variables and the broad-scale MEM submodel, are *lower* than in Fig. 3a (the fraction identifiers [a] to [h] are the same as in Fig. 2a). The total amount of explained variation remains the same, however, to within rounding errors of the numbers displayed in the figure.

Fractions [a], [c], and [f] are unaffected by the hierarchical approach since they involve only the pure contributions of the environmental and fine-scale MEM matrices and their jointly explained variation. The hierarchy affects the broad-scale MEM submodel in that this submodel now completely incorporates fractions [e] and [g] that, otherwise, would be common to the broad- and fine-scale MEM submodels. In the hierarchical partitioning result, the  $R_a^2$  of the broad-scale MEM submodel is perfectly represented in the partition result.

To make the numerical example complete, Fig. 3c illustrates the result of the proportional apportioning method. The intersection fractions are apportioned proportionally to the variation explained by each submodel, that is, fraction [b + d] of Fig. 2a for the broad-scale submodel and fraction [c + f] for the fine-scale submodel. The numerical artifacts represented by fractions [e] and [g] were not taken into account in calculating the relative importance of the submodels because they are the quantities that must be apportioned between [b + d] and [c + f].



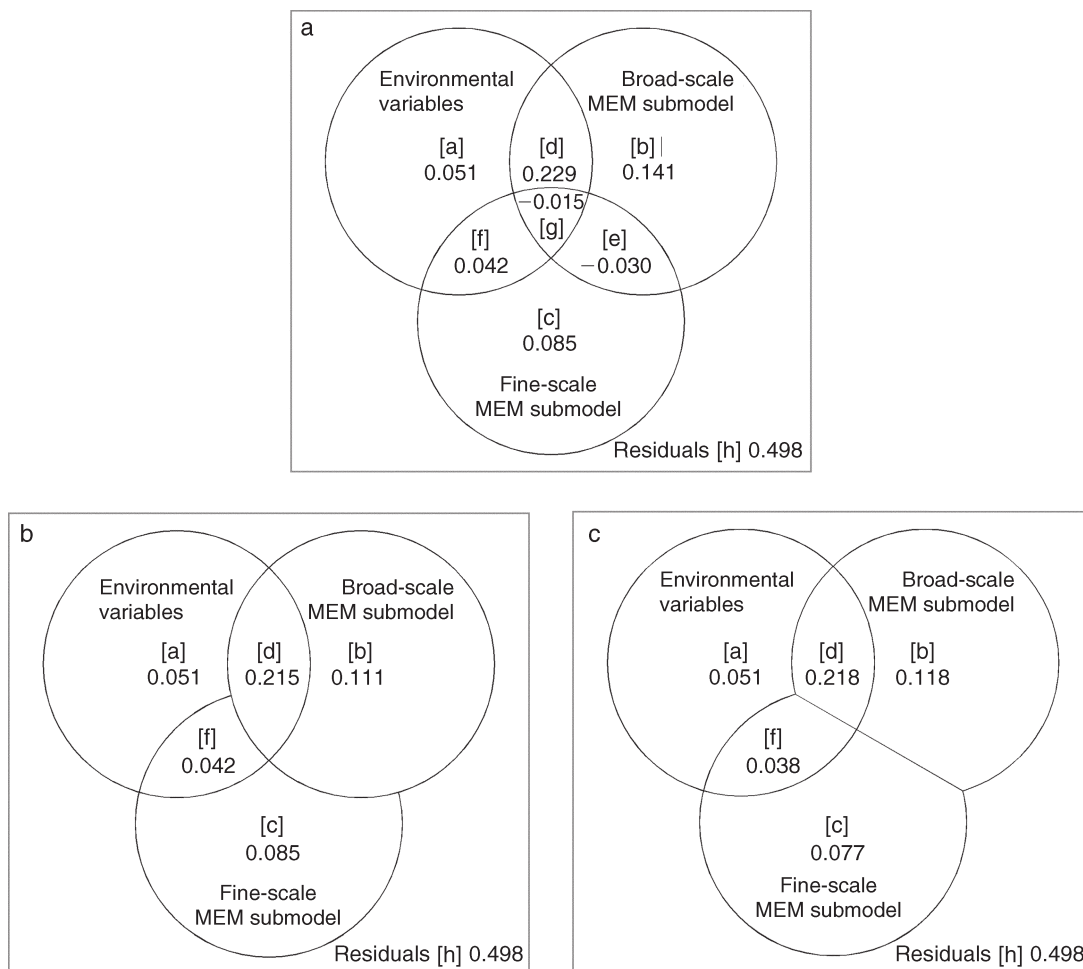


FIG. 3. Venn diagrams of the partitioning results for the oribatid mite data: (a) classical, nonhierarchical partitioning; (b) hierarchical partitioning of the fraction shared by the two MEM submodels, taking into account the priority of the broad-scale submodel over the fine-scale submodel; (c) proportional apportioning.

### CONCLUSION

The development described in this paper sheds a new light on the interpretation of spatial modeling results implying explanatory environmental variables as well as submodels of spatial eigenfunctions involving two or more scales. Ezekiel's adjustment of the  $R^2$  statistic yields an unbiased estimate of the explained variance, but when used in the context of variation partitioning it has the side effect of generating spurious nonzero joint fractions between orthogonal explanatory matrices. Two solutions are proposed, and the choice between them requires users to decide first if the processes underlying the spatial fractions are of a hierarchical nature or not. The nonhierarchical solution translates into the proportional apportioning of the joint fractions to the spatial submodels. Invoking hierarchy theory to put the two or more spatial submodels in a logical order of importance allows one to ascribe each of these fractions to one and

only one spatial scale. This clarifies the interpretation of the spatial components of the response data variation.

The common fractions resulting from the analysis of environmental explanatory matrices are not involved in this hierarchical operation because these matrices are not a priori linearly independent of one another or of the spatial eigenfunction submodels. For explanatory environmental variables, Peres-Neto et al. (2006) have shown that partitioning based on adjusted  $R^2$  statistics ( $R_a^2$ ) produces unbiased estimates of all fractions of variation: unique (e.g., [a] and [c] in the two explanatory matrices case) and shared (e.g., [b]). The partitioning of univariate or multivariate response data, e.g., community composition data, between environmental explanatory variables and spatial submodels is of great interest since it allows ecologists to discover structures in the response data that are explainable by both the environmental variables and spatial models at different scales.

Some points in linear variation partitioning remain to be investigated. The use of categorical variables in variation partitioning is still a subject of discussion. Random categorical variables such as the type of soil observed at sites may be treated as a set of binary variables and included as such in an environmental matrix, but fixed ANOVA factors (e.g., in a manipulative field ecological experiment) are clearly of a different nature, and Ezekiel's adjustment should not be applied to them. Indeed, Healy (1984) pointed out that Ezekiel's (1930) adjusted  $R^2$  makes sense and should be used when explanatory matrix  $\mathbf{X}$  contains observed values of random variables, but not for ANOVA factors. Therefore, the procedure developed in this Note should not be applied to cases involving fixed ANOVA factors, only to orthogonal models of spatial eigenfunctions. Although they are called "spatial," eigenfunction models can also be used for multiscale analysis of time series.

#### ACKNOWLEDGMENTS

We are grateful to the reviewers for interesting discussion. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grant no. 7738 to P. Legendre.

#### LITERATURE CITED

- Allen, T. F. H., and T. B. Starr. 1982. *Hierarchy—Perspectives for ecological complexity*. University of Chicago Press, Chicago, Illinois, USA.
- Blanchet, F. G., P. Legendre, and D. Borcard. 2008a. Modelling directional spatial processes in ecological data. *Ecological Modelling* 215:325–336.
- Blanchet, F. G., P. Legendre, and D. Borcard. 2008b. Forward selection of explanatory variables. *Ecology* 89:2623–2632.
- Blanchet, F. G., P. Legendre, and D. Borcard. 2009. Erratum to "Modelling directional spatial processes in ecological data" [Ecological Modelling 215 (2008): 325–336]. *Ecological Modelling* 220:82–83.
- Borcard, D., F. Gillet, and P. Legendre. 2011. *Numerical ecology with R*. Use R! series. Springer Science, New York, New York, USA.
- Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153:51–68.
- Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structures of ecological data at all scales. *Ecology* 85:1826–1832.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73:1045–1055.
- Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196:483–493.
- Ezekiel, M. 1930. *Methods of correlation analysis*. John Wiley, New York, New York, USA.
- Healy, M. J. R. 1984. The use of  $R^2$  as a measure of goodness of fit. *Journal of the Royal Statistical Society Series A* 147:608–609.
- Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271–280.
- Legendre, P., and L. Legendre. 2012. *Numerical ecology*, Third English edition. Elsevier Science BV, Amsterdam, The Netherlands.
- Ohtani, K. 2000. Bootstrapping  $R^2$  and adjusted  $R^2$  in regression analysis. *Economic Modelling* 17:473–483.
- O'Neill, R. V., R. H. Gardner, B. T. Milne, M. G. Turner, and B. Jackson. 1991. Heterogeneity and spatial hierarchies. Pages 85–96 in J. Kolasa and S. T. A. Pickett, editors. *Ecological heterogeneity*. Springer-Verlag, New York, New York, USA.
- Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87:2614–2625.
- Simon, H. A. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* 106:467–482.

#### SUPPLEMENTAL MATERIAL

##### Appendix

Variation partitioning formulas, hierarchical and proportional methods (*Ecological Archives* E093-106-A1).

##### Supplement

Functions in R language to compute hierarchical and proportional variation partitioning for eigenfunction submodels (*Ecological Archives* E093-106-S1).

**Legendre, P., D. Borcard and D. W. Roberts. 2012. Variation partitioning involving orthogonal spatial eigenfunction submodels. *Ecology* 93: 1234–1240.**

Appendix A. Variation partitioning formulas, hierarchical and proportional methods

This Appendix presents formulas for the calculation of fractions following the hierarchical and proportional variation partitioning methods for eigenfunction submodels (MEM or AEM). Classical variation partitioning (Borcard et al. 1992, Peres-Neto et al. 2006) only involves matrices of random variables. Different situations, involving different combinations of matrices of random (e.g. environmental) variables and spatial eigenfunction submodels, are described in Tables A1 to A3.

Venn diagrams illustrating the results of proportional apportioning of the shared fractions in variation partitioning are presented in Figs. A1 to A3. Figures illustrating the results of hierarchical partitioning of the shared fractions for two and three explanatory matrices are shown in the main paper, Figs. 1 and 2, and in Fig. A4 for four explanatory matrices.



Table A1. Variation partitioning for two explanatory matrices containing eigenfunction submodels. Fractions identifiers: see Figs. 1 and 2a. *Notation:*  $\emptyset$  is the empty set; in the present table, is denotes the absence of a value. Indications about the logic involved in the computation of some fractions follow the sign #.

---

Hierarchical partitioning of the shared fraction

$$[a] = [a] + [b]$$

$$[b] = \emptyset \quad \# \text{ Transferred to } [a]$$

$$[c] = [c]$$

$$[d] = [d] \quad \# \text{ Residual fraction}$$

Proportional apportioning of the shared fraction

$$[a] = [a] + ([b] \times [a] / ([a] + [c]))$$

$$[b] = \emptyset \quad \# \text{ Transferred to } [a] \text{ and } [c]$$

$$[c] = [c] + ([b] \times [c] / ([a] + [c]))$$

$$[d] = [d] \quad \# \text{ Residual fraction}$$


---

Two matrices of MEM

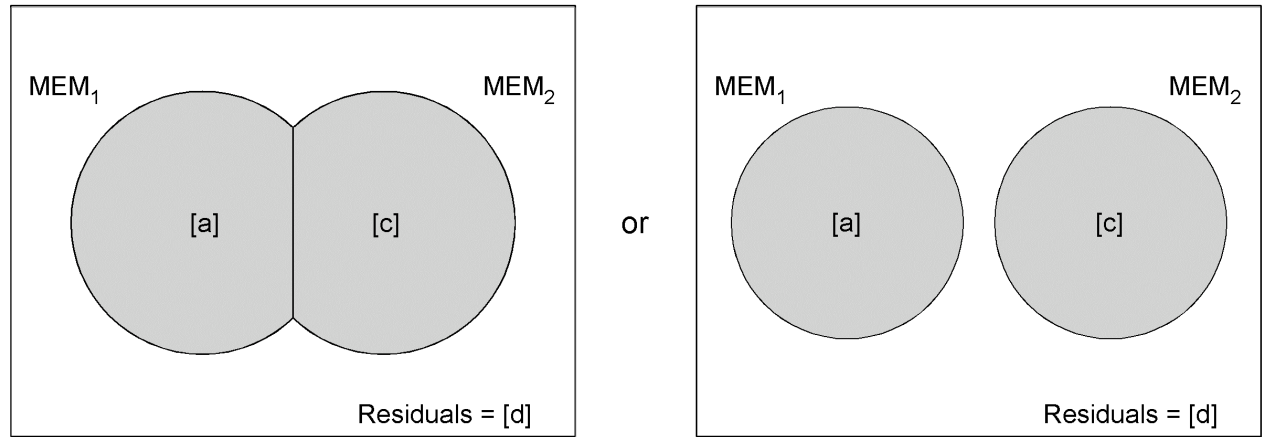


Fig. A1. Venn diagrams illustrating the result of proportional apportioning of the shared fractions in variation partitioning involving two matrices of spatial eigenfunctions (MEM). Alternative graphical solutions are proposed. In gray, the fractions of variation explained by MEM.

Table A2. Variation partitioning in the 3-matrix case. Matrix 1 may contain random explanatory variables. (a) The last two, or (b) all three matrices, contain eigenfunction submodels. Fractions identifiers: see Fig. 3a. Notation as in Table A1.

(a) Matrix 1 contains random explanatory variables, matrices 2 and 3 contain eigenfunction submodels

---

Hierarchical partitioning of the shared fractions (Fig. 3b)

$[a] = [a]$	
$[b] = [b] + [e]$	
$[c] = [c]$	
$[d] = [d] + [g]$	
$[e] = \emptyset$	# Transferred to [b]
$[f] = [f]$	
$[g] = \emptyset$	# Transferred to [d]
$[h] = [h]$	# Residual fraction

---

Proportional apportioning of the shared fractions

$e.to.b = [e] \times (([b] + [d]) / ([b] + [d] + [c] + [f]))$	# Portion of [e] transferred to [b]
$e.to.c = [e] \times (([c] + [f]) / ([b] + [d] + [c] + [f]))$	# Portion of [e] transferred to [c]
$g.to.d = [g] \times (([b] + [d]) / ([b] + [d] + [c] + [f]))$	# Portion of [g] transferred to [d]
$g.to.f = [g] \times (([c] + [f]) / ([b] + [d] + [c] + [f]))$	# Portion of [g] transferred to [f]
$[a] = [a]$	
$[b] = [b] + e.to.b$	
$[c] = [c] + e.to.c$	
$[d] = [d] + g.to.d$	
$[e] = \emptyset$	# Transferred to [b] and [c]
$[f] = [f] + g.to.f$	
$[g] = \emptyset$	# Transferred to [d] and [f]
$[h] = [h]$	# Residual fraction

---

## (b) Three matrices of eigenfunction submodels

---

Hierarchical partitioning of the shared fractions (Fig. 3c)

$[a] = [a] + [d] + [f] + [g]$	
$[b] = [b] + [e]$	
$[c] = [c]$	
$[d] = \emptyset$	# Transferred to [a]
$[e] = \emptyset$	# Transferred to [b]
$[f] = \emptyset$	# Transferred to [a]
$[g] = \emptyset$	# Transferred to [a]
$[h] = [h]$	# Residual fraction

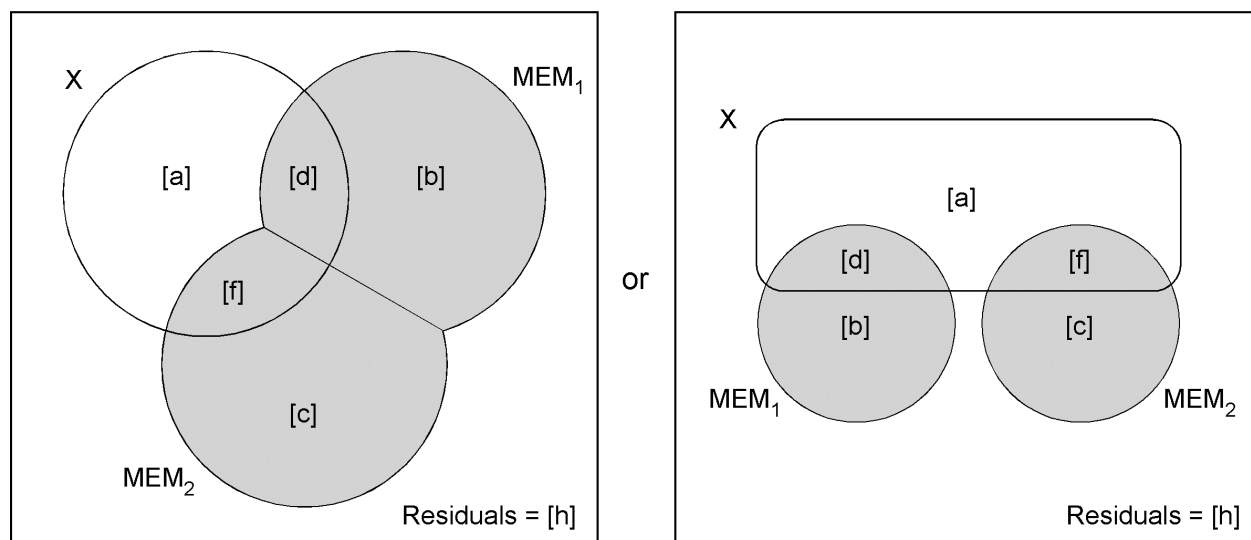
---

Proportional apportioning of the shared fractions

$d.to.a = [d] \times ([a] / ([a] + [b]))$	# Portion of [d] transferred to [a]
$d.to.b = [d] \times ([b] / ([a] + [b]))$	# Portion of [d] transferred to [b]
$e.to.b = [e] \times ([b] / ([b] + [c]))$	# Portion of [e] transferred to [b]
$e.to.c = [e] \times ([c] / ([b] + [c]))$	# Portion of [e] transferred to [c]
$f.to.a = [f] \times ([a] / ([a] + [c]))$	# Portion of [f] transferred to [a]
$f.to.c = [f] \times ([c] / ([a] + [c]))$	# Portion of [f] transferred to [c]
$g.to.a = [g] \times ([a] / ([a] + [b] + [c]))$	# Portion of [g] transferred to [a]
$g.to.b = [g] \times ([b] / ([a] + [b] + [c]))$	# Portion of [g] transferred to [b]
$g.to.c = [g] \times ([c] / ([a] + [b] + [c]))$	# Portion of [g] transferred to [c]
$[a] = [a] + d.to.a + f.to.a + g.to.a$	
$[b] = [b] + d.to.b + e.to.b + g.to.b$	
$[c] = [c] + e.to.c + f.to.c + g.to.c$	
$[d] = \emptyset$	# Transferred to [a] and [b]
$[e] = \emptyset$	# Transferred to [b] and [c]
$[f] = \emptyset$	# Transferred to [a] and [c]
$[g] = \emptyset$	# Transferred to [a], [b] and [c]
$[h] = [h]$	# Residual fraction

---

(a) One matrix of random variables  $X$ ,  
two matrices of MEM



(b) Three matrices of MEM

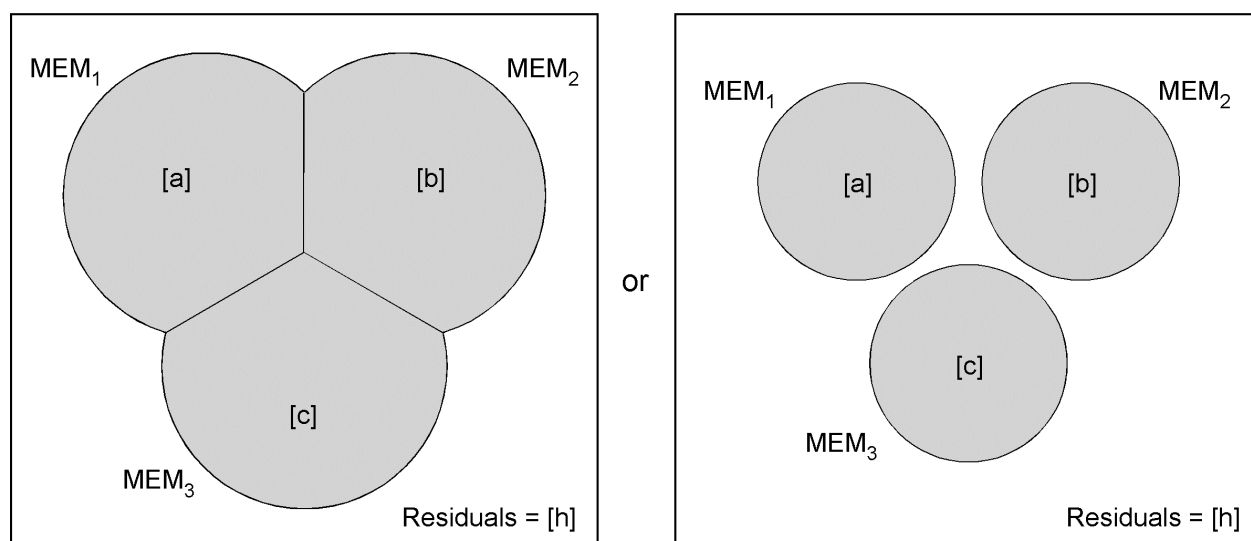


Fig. A2. Venn diagrams illustrating the result of proportional apportioning of the shared fractions in variation partitioning involving (a) one matrix of random (e.g. environmental) variables and two matrices of spatial eigenfunctions (MEM), or (b) three matrices of spatial eigenfunctions (MEM). Alternative graphical solutions are proposed in each case. The fractions of variation explained by MEM matrices are in gray.

Table A3. Variation partitioning in the 4-matrix case, for matrices of random explanatory variables or matrices of eigenfunctions. Fractions identifiers: see Fig. A4a. Notation as in Table A1.

(a) Matrices 1 and 2 contain random explanatory variables, matrices 3 and 4 contain eigenfunction submodels

---

Hierarchical partitioning of the shared fractions (Fig. A4b)

$[a] = [a]$	
$[b] = [b]$	
$[c] = [c] + [j]$	
$[d] = [d]$	
$[e] = [e]$	
$[f] = [f] + [m]$	
$[g] = [g] + [n]$	
$[h] = [h]$	
$[i] = [i]$	
$[k] = [k]$	
$[l] = [l] + [o]$	
$[j] = \emptyset$	# Transferred to [c]
$[m] = \emptyset$	# Transferred to [f]
$[n] = \emptyset$	# Transferred to [g]
$[o] = \emptyset$	# Transferred to [l]
$[p] = [p]$	# Residual fraction

---

Proportional apportioning of the shared fractions (Fig. A3a)

$$\begin{aligned}
 j.\text{to}.c &= [j] \times [c+f+g+l] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [j] \text{ transferred to } [c] \\
 j.\text{to}.d &= [j] \times [d+h+i+k] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [j] \text{ transferred to } [d] \\
 m.\text{to}.f &= [m] \times [c+f+g+l] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [m] \text{ transferred to } [f] \\
 m.\text{to}.i &= [m] \times [d+h+i+k] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [m] \text{ transferred to } [i] \\
 n.\text{to}.g &= [n] \times [c+f+g+l] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [n] \text{ transferred to } [g] \\
 n.\text{to}.h &= [n] \times [d+h+i+k] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [n] \text{ transferred to } [h] \\
 o.\text{to}.k &= [o] \times [d+h+i+k] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [o] \text{ transferred to } [k] \\
 o.\text{to}.l &= [o] \times [c+f+g+l] / ([c+f+g+l]+[d+h+i+k]) && \# \text{ Portion of } [o] \text{ transferred to } [l] \\
 [c] &= [c] + j.\text{to}.c \\
 [d] &= [d] + j.\text{to}.d \\
 [f] &= [f] + m.\text{to}.f \\
 [g] &= [g] + n.\text{to}.g \\
 [h] &= [h] + n.\text{to}.h \\
 [i] &= [i] + m.\text{to}.i \\
 [k] &= [k] + o.\text{to}.k \\
 [l] &= [l] + o.\text{to}.l \\
 [j] &= \emptyset && \# \text{ Transferred to } [c] \text{ and } [d] \\
 [m] &= \emptyset && \# \text{ Transferred to } [f] \text{ and } [i] \\
 [n] &= \emptyset && \# \text{ Transferred to } [g] \text{ and } [h] \\
 [o] &= \emptyset && \# \text{ Transferred to } [k] \text{ and } [l] \\
 [p] &= [p] && \# \text{ Residual fraction}
 \end{aligned}$$


---



(b) Matrix 1 contains random explanatory variables, matrices 2, 3 and 4 contain eigenfunction submodels

---

Hierarchical partitioning of the shared fractions (Fig. A4c)

$$[a] = [a]$$

$$[b] = [b] + [f] + [i] + [m]$$

$$[c] = [c] + [j]$$

$$[d] = [d]$$

$$[e] = [e] + [k] + [l] + [o]$$

$$[g] = [g] + [n]$$

$$[h] = [h]$$

$$[f] = [i] = [j] = [k] = [l] = [m] = [n] = [o] = \emptyset \quad \# \text{ Transferred to other fractions}$$

$$[p] = [p]$$


---

Proportional apportioning of the shared fractions (Fig. A3b)

$f.to.b = [f] \times [b+e] / ([b+e]+[c+g])$	# Portion of [f] transferred to [b]
$f.to.c = [f] \times [c+g] / ([b+e]+[c+g])$	# Portion of [f] transferred to [c]
$i.to.b = [i] \times [b+e] / ([b+e]+[d+h])$	# Portion of [i] transferred to [b]
$i.to.d = [i] \times [d+h] / ([b+e]+[d+h])$	# Portion of [i] transferred to [d]
$j.to.c = [j] \times [c+g] / ([c+g]+[d+h])$	# Portion of [j] transferred to [c]
$j.to.d = [j] \times [d+h] / ([c+g]+[d+h])$	# Portion of [j] transferred to [d]
$k.to.e = [k] \times [b+e] / ([b+e]+[d+h])$	# Portion of [k] transferred to [e]
$k.to.h = [k] \times [d+h] / ([b+e]+[d+h])$	# Portion of [k] transferred to [h]
$l.to.e = [l] \times [b+e] / ([b+e]+[c+g])$	# Portion of [l] transferred to [e]
$l.to.g = [l] \times [c+g] / ([b+e]+[c+g])$	# Portion of [l] transferred to [g]
$m.to.b = [m] \times [b+e] / ([b+e]+[c+g]+[d+h])$	# Portion of [m] transferred to [b]
$m.to.c = [m] \times [c+g] / ([b+e]+[c+g]+[d+h])$	# Portion of [m] transferred to [c]
$m.to.d = [m] \times [d+h] / ([b+e]+[c+g]+[d+h])$	# Portion of [m] transferred to [d]
$n.to.g = [n] \times [c+g] / ([c+g]+[d+h])$	# Portion of [n] transferred to [g]
$n.to.h = [n] \times [d+h] / ([c+g]+[d+h])$	# Portion of [n] transferred to [h]
$o.to.e = [o] \times [b+e] / ([b+e]+[c+g]+[d+h])$	# Portion of [o] transferred to [e]
$o.to.g = [o] \times [c+g] / ([b+e]+[c+g]+[d+h])$	# Portion of [o] transferred to [g]
$o.to.h = [o] \times [d+h] / ([b+e]+[c+g]+[d+h])$	# Portion of [o] transferred to [h]
$[b] = [b] + f.to.b + m.to.b + i.to.b$	
$[c] = [c] + j.to.c + f.to.c + m.to.c$	
$[d] = [d] + j.to.d + i.to.d + m.to.d$	
$[e] = [e] + k.to.e + l.to.e + o.to.e$	
$[g] = [g] + l.to.g + n.to.g + o.to.g$	
$[h] = [h] + k.to.h + n.to.h + o.to.h$	
$[f] = [i] = [j] = [k] = [l] = [m] = [n] = [o] = \emptyset$	# Transferred to other fractions
$[p] = [p]$	# Residual fraction

---

## (c) Four matrices of eigenfunction submodels

Hierarchical partitioning of the shared fractions (Fig. A4d)

$$[a] = [a] + [e] + [g] + [h] + [k] + [l] + [n] + [o]$$

$$[b] = [b] + [f] + [i] + [m]$$

$$[c] = [c] + [j]$$

$$[d] = [d]$$

$$[e] = [f] = [g] = [h] = [i] = [j] = [k] = [l] = [m] = [n] = [o] = \emptyset \quad \# \text{ Transferred to other fractions}$$

$$[p] = [p] \quad \# \text{ Residual fraction}$$

Proportional apportioning of the shared fractions (Fig. A3c)

$$e.to.a = [e] \times [a] / ([a] + [b]) \quad \# \text{ Portion of } [e] \text{ transferred to } [a]$$

$$e.to.b = [e] \times [b] / ([a] + [b]) \quad \# \text{ Portion of } [e] \text{ transferred to } [b]$$

$$f.to.b = [f] \times [b] / ([b] + [c]) \quad \# \text{ Portion of } [f] \text{ transferred to } [b]$$

$$f.to.c = [f] \times [c] / ([b] + [c]) \quad \# \text{ Portion of } [f] \text{ transferred to } [c]$$

$$g.to.a = [g] \times [a] / ([a] + [c]) \quad \# \text{ Portion of } [g] \text{ transferred to } [a]$$

$$g.to.c = [g] \times [c] / ([a] + [c]) \quad \# \text{ Portion of } [g] \text{ transferred to } [c]$$

$$h.to.a = [h] \times [a] / ([a] + [d]) \quad \# \text{ Portion of } [h] \text{ transferred to } [a]$$

$$h.to.d = [h] \times [d] / ([a] + [d]) \quad \# \text{ Portion of } [h] \text{ transferred to } [d]$$

$$i.to.b = [i] \times [b] / ([b] + [d]) \quad \# \text{ Portion of } [i] \text{ transferred to } [b]$$

$$i.to.d = [i] \times [d] / ([b] + [d]) \quad \# \text{ Portion of } [i] \text{ transferred to } [d]$$

$$j.to.c = [j] \times [c] / ([c] + [d]) \quad \# \text{ Portion of } [j] \text{ transferred to } [c]$$

$$j.to.d = [j] \times [d] / ([c] + [d]) \quad \# \text{ Portion of } [j] \text{ transferred to } [d]$$

$$k.to.a = [k] \times [a] / ([a] + [b] + [d]) \quad \# \text{ Portion of } [k] \text{ transferred to } [a]$$

$$k.to.b = [k] \times [b] / ([a] + [b] + [d]) \quad \# \text{ Portion of } [k] \text{ transferred to } [b]$$

$$k.to.d = [k] \times [d] / ([a] + [b] + [d]) \quad \# \text{ Portion of } [k] \text{ transferred to } [d]$$

$$l.to.a = [l] \times [a] / ([a] + [b] + [c]) \quad \# \text{ Portion of } [l] \text{ transferred to } [a]$$

$$l.to.b = [l] \times [b] / ([a] + [b] + [c]) \quad \# \text{ Portion of } [l] \text{ transferred to } [b]$$

$$l.to.c = [l] \times [c] / ([a] + [b] + [c]) \quad \# \text{ Portion of } [l] \text{ transferred to } [c]$$

$$m.to.b = [m] \times [b] / ([b] + [c] + [d]) \quad \# \text{ Portion of } [m] \text{ transferred to } [b]$$

$$m.to.c = [m] \times [c] / ([b] + [c] + [d]) \quad \# \text{ Portion of } [m] \text{ transferred to } [c]$$

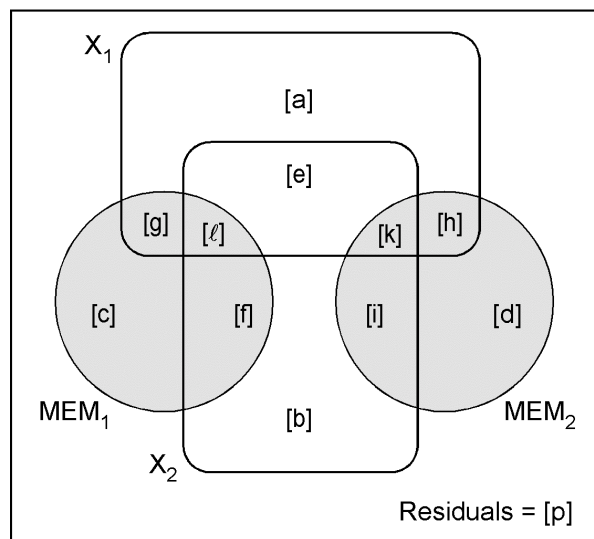
$$m.to.d = [m] \times [d] / ([b] + [c] + [d]) \quad \# \text{ Portion of } [m] \text{ transferred to } [d]$$

$$n.to.a = [n] \times [a] / ([a] + [c] + [d]) \quad \# \text{ Portion of } [n] \text{ transferred to } [a]$$

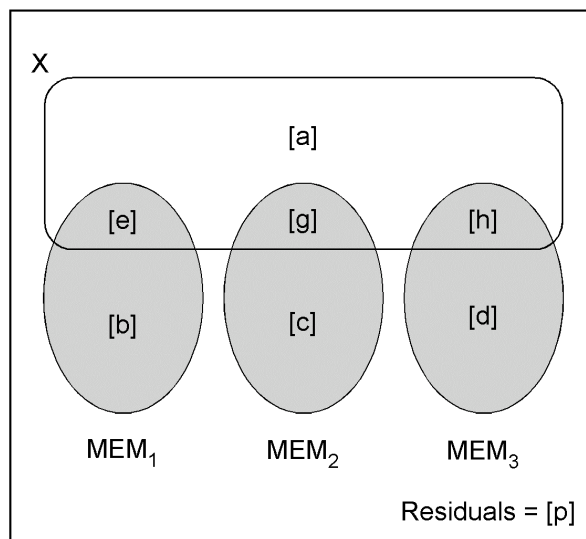
$$\begin{aligned}
n.to.c &= [n] \times [c] / ([a]+[c]+[d]) && \# \text{ Portion of } [n] \text{ transferred to } [c] \\
n.to.d &= [n] \times [d] / ([a]+[c]+[d]) && \# \text{ Portion of } [n] \text{ transferred to } [d] \\
o.to.a &= [o] \times [a] / ([a]+[b]+[c]+[d]) && \# \text{ Portion of } [o] \text{ transferred to } [a] \\
o.to.b &= [o] \times [b] / ([a]+[b]+[c]+[d]) && \# \text{ Portion of } [o] \text{ transferred to } [b] \\
o.to.c &= [o] \times [c] / ([a]+[b]+[c]+[d]) && \# \text{ Portion of } [o] \text{ transferred to } [c] \\
o.to.d &= [o] \times [d] / ([a]+[b]+[c]+[d]) && \# \text{ Portion of } [o] \text{ transferred to } [d] \\
[a] &= [a] + e.to.a + g.to.a + h.to.a + k.to.a + l.to.a + n.to.a + o.to.a \\
[b] &= [b] + e.to.b + f.to.b + i.to.b + k.to.b + l.to.b + m.to.b + o.to.b \\
[c] &= [c] + f.to.c + g.to.c + j.to.c + l.to.c + m.to.c + n.to.c + o.to.c \\
[d] &= [d] + h.to.d + i.to.d + j.to.d + k.to.d + m.to.d + n.to.d + o.to.d \\
[e] &= [f] = [g] = [h] = [i] = [j] = [k] = [l] = [m] = [n] = [o] = \emptyset && \# \text{ Transferred to other fractions} \\
[p] &= [p] && \# \text{ Residual fraction}
\end{aligned}$$


---

(a) Two matrices of random variables  $X$ , two matrices of MEM



(b) One matrix of random variables  $X$ , three matrices of MEM



(c) Four matrices of MEM

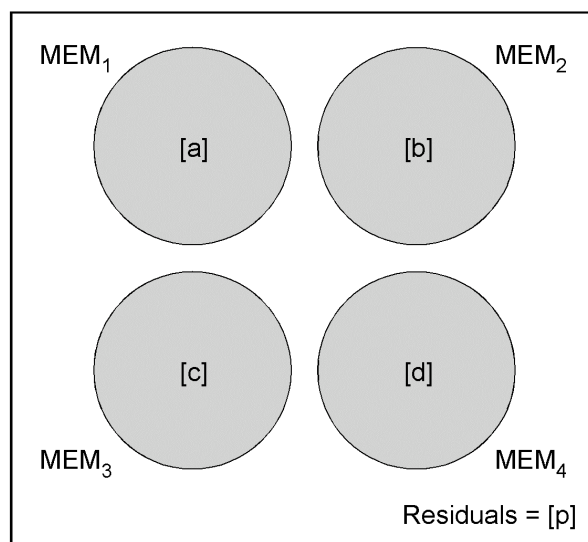


Fig. A3. Venn diagrams illustrating the result of proportional apportioning of the shared fractions in variation partitioning involving (a) two matrices of random (e.g. environmental) variables and two matrices of spatial eigenfunctions (MEM), (b) one matrix of random (e.g. environmental) variables and three matrices of spatial eigenfunctions (MEM), or (c) four matrices of spatial eigenfunctions (MEM). The fractions of variation explained by MEM matrices are in gray.

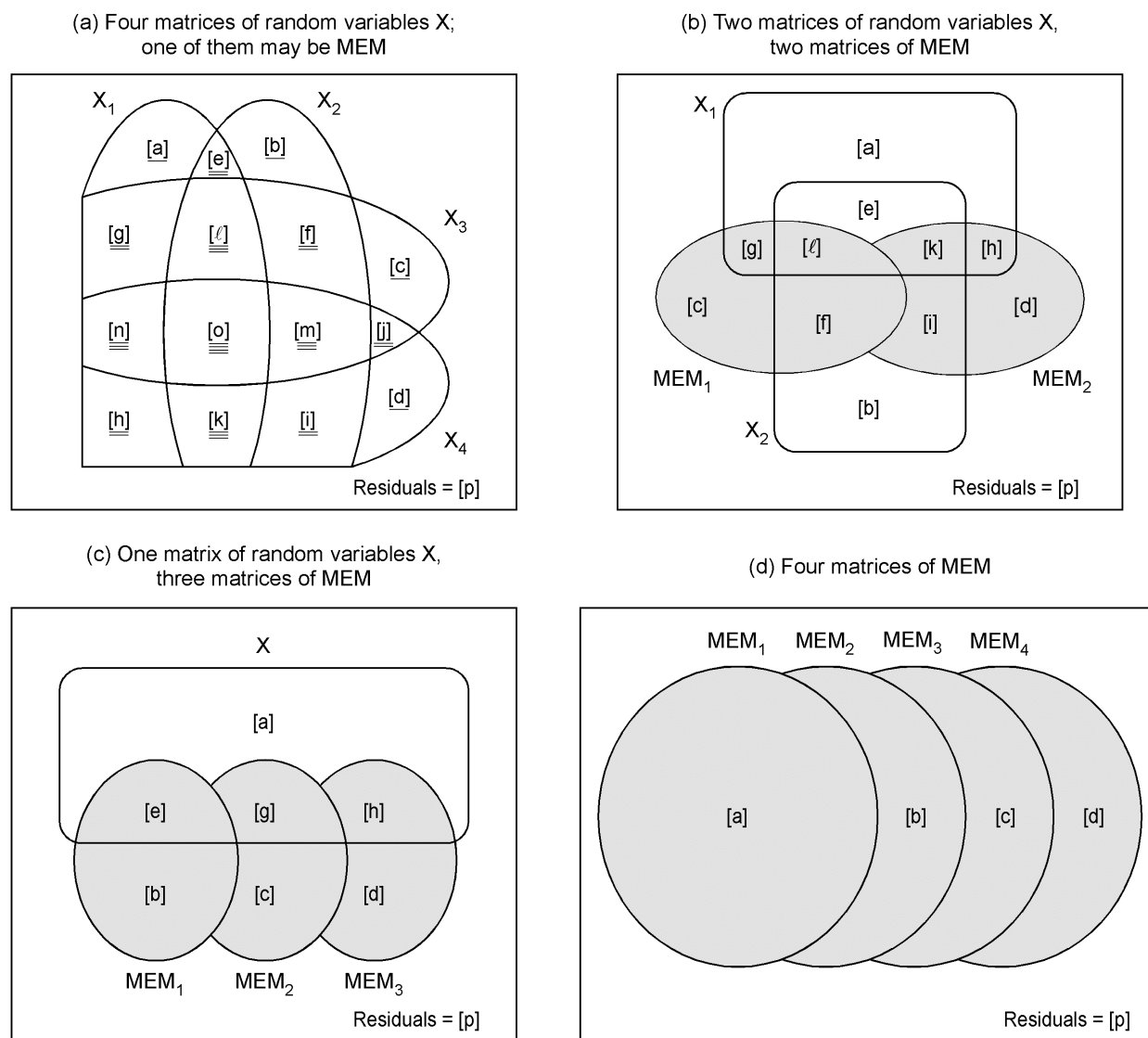


Fig. A4. Venn diagrams of variation partitioning results for four explanatory matrices. (a) Classical partitioning; the number of lines underneath the fraction identifiers corresponds to the number of explanatory matrices  $X$  they intersect. (b, c, d) Result of hierarchical partitioning.