

A Suite of Matlab functions to compare methods of variation partitioning in species data matrices.

Pedro R. Peres-Neto¹ and Pierre Legendre²

May, 2005

¹ Department of Biology
University of Regina
Regina, SK S4S 0A2, Canada

Email: Pedro.Peres-Neto@uregina.ca

² Département de sciences biologiques
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal, Québec H3C 3J7, Canada

Email: Pierre.Legendre@umontreal.ca

Description

In this zip file we offer the suite of Matlab functions used in our simulation study to compare the raw data and distance-based methods when partitioning the variation in species data matrices.

Each species data set was analyzed using two methods of variation partitioning against a set of environmental variables, described in the data generation section, and a set of spatial variables created as follows. For canonical partitioning, we used (1) the X and Y geographic coordinates of the 100 points forming the sampling grid, (2) a polynomial function of order 3 of the X and Y geographic coordinates, and (3) a set of PCNM variables (57) selected by forward selection procedure similar to that of the program Canoco. This resulted in a selection of 4.5 PCNM variables, on average, per data set (range: from 0 to 15). For partitioning on distance matrices, we used (1) a geographic distance matrix $D(XY)$ computed from the X and Y geographic coordinates, (2) a distance matrix $D(\text{polyn.})$ obtained by computing Euclidean distances based on the 3rd-order polynomial of the geographic coordinates described in the previous paragraph, and (3) a distance matrix $\ln(D(XY))$ obtained by taking the natural logarithm of the values in $D(XY)$.

This zip file contains 15 files: this guide file (2 files: pdf and text), 8 Matlab functions and 5 text files containing the necessary files to run one of the scenarios considered in our study (i.e., Table 1, scenario D) as follows. Each Matlab function contains important details regarding their use and implementation. The main Matlab file that executes the simulation is "ExampleSimulationFile.m". Since all functions can be used independently, they can be easily applied to any data set of interest. Please feel free to contact us with questions and/or suggestions.

List of files:

1 - DistanceVariationPartition.m: performs a variation partitioning on a distance matrix (or similarity) Y based on two data distance matrices X and W. Tests of significance on fractions are performed by permutation.

2 - eucl.m: Calculates the Euclidean distances among a set of points, or between a reference point and a set of points, or among all possible pairs of two sets of points, in P dimensions. Returns a single distance for two points. (author: Richard Strauss).

3 - ExampleSimulationFile.m: one of the files used to run a complete set of simulations. The files used in this example (Envir025.txt and Species025.txt) relates to Scenario D in Table 1 in Legendre, P., D. Borcard and P. R. Peres-Neto (Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecological Monographs, 2005). Envir025.txt and Species025.txt were generated by the accompanying program in this supplement: SimSSD4.exe using the following parameters: 1000 100 100 2 5 0.50 -9 0 0 0 15 15 1 2 100 5 5

4 - ForwardSelectionRDA.m: performs a forward selection procedure of regressors in RDA. Function follows implementation of ter Braak and Smilauer with the exception that the procedure stops once a regressor is considered non-significant according to the alpha level established. Note that if an alpha=1 is set, then the procedure gives the significance level for all variables as in Canoco.

Reference: ter Braak, C. J. F., and P. Smilauer. 2002. Canoco reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). Microcomputer Power, Ithaca, New York.

5 - HellingerTransformation.m: performs a HellingerTransformation (Legendre and Gallagher 2001).

Reference: Legendre, P. and E. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271-280.

6 - RDATest.m: performs a permutation test for assessing the relationship between two multivariate data matrices Y and X based on redundancy analysis (RDA).

7 - TransformIntoBinary.m: Transforms a species data matrix based on abundance into a presence/absence (binary) species data matrix.

8 - VariationPartitionRDA.m: performs a variation partitioning on a data table Y based on two data matrices X and W based on redundancy analysis (RDA). Tests of significance on fractions are performed by permutation.

9 - Coord.txt: contains the X and Y coordinates of the points selected by the sampling design generated by the accompanying program in this supplement: SimSSD4.exe.

10 - Envir025.txt: contains successive tables (1000) of generated environmental data generated by the accompanying program in this supplement: SimSSD4.exe (see ExampleSimulationFile.m above for details regarding simulation parameters).

11 - pcnm.txt: contains the 57 PCNM variables used in all the simulations.

12 - polynomial.txt: contains polynomial function of order 3 of the X and Y geographic coordinates.

13 - Species025.txt: contains successive tables (1000) of generated species data generated by the accompanying program in this supplement: SimSSD4.exe (see ExampleSimulationFile.m above for details regarding simulation parameters).