

# R Package

## Spatial analysis exercise

Pierre Legendre

April 1992 / February 1994 / October 1996 / May 1999

The exercise proposed here concerns data from a sampling campaign in the Thau lagoon in southern France. These data will allow you to compute correlograms (Moran and Geary statistics), simple and partial Mantel tests, and a Mantel correlogram. The data are succinctly described in the next paragraph.

During an ecological research program on a brackish lagoon of southern France (Amanieu *et al.*, 1989), sampling has been carried out at 20 stations (depth: 0.5 m), on October 25, 1988. The selection process that led to choosing these particular 20 stations is described in Legendre *et al.* (1989). The following Table presents some of the measured variables; they are also found in file “Thau data (20x7)” where columns are separated by tabulators.

Station no	Phytoplankton <sup>1</sup> ln(x + 1)	NH <sub>4</sub> <sup>2</sup> (ln)	Phaeopigments (ln)	DNA <sup>3</sup> (ln)	Geogr. coordinates <sup>4</sup> X (km)    Y (km)	
1	0.31	0.307	0.184	0.274	3	10
2	-0.31	0.207	0.212	0.213	5	9
3	-0.38	0.140	0.229	0.134	6	8
4	-0.89	1.371	0.287	0.177	6	10
5	-1.09	1.447	0.242	0.091	8	9
6	-0.43	0.668	0.531	0.272	9	10
7	0.49	0.300	0.948	0.460	10	7
8	0.89	0.329	1.389	0.253	11	6
9	0.41	0.207	0.765	0.235	12	5
10	0.23	0.223	0.737	0.362	12	7
11	-0.23	0.788	0.454	0.824	12	9
12	-0.06	1.112	0.395	0.419	13	8
13	-0.51	1.273	0.247	0.398	15	2
14	0.47	0.956	0.449	0.172	15	4
15	0.34	0.708	0.457	0.141	15	5
16	0.10	0.637	0.386	0.360	16	1
17	0.23	0.519	0.481	0.261	16	2
18	0.48	0.247	0.468	0.450	16	4
19	-0.11	1.664	0.321	0.287	17	6
20	0.06	0.182	0.380	0.510	18	4

<sup>1</sup> Phytoplankton biomass is estimated by chlorophyll *a* density, expressed in mg·L<sup>-1</sup>. This variable has been normalised using a log transformation [ $x' = \ln(x + 1)$ ]. It is found in file “Chla (20x1)”.

<sup>2</sup> Variable NH<sub>4</sub>, expressed in µmol·L<sup>-1</sup>, has been normalised using a log transformation [ $x' = \ln(x)$ ].

<sup>3</sup> Bacterial DNA was measured by the incorporation of tritiated thymidine by cells in culture.

<sup>4</sup> The geographic positions of the sampling stations are expressed by X and Y coordinates (measured in km) in a Cartesian reference plane with arbitrary zero position. They are found in file “Thau XY (20x2)”.

## Data

For the exercise, you can use either the raw data table above, or the corresponding ASCII data files available in the present folder. A single matrix containing all the data is available on the Web site.

- 1) BNA,MA (20x2): ASCII file containing two bacterial variables: BNA = aerobic heterotrophic bacteria growing on *Nutrient agar* medium (low NaCl concentration), and MA = aerobic heterotrophic bacteria growing on *Marine agar* (34 g·L<sup>-1</sup> NaCl).
- 2) Chla (20x1): ASCII file with variable Chlorophyll *a* only, transformed by eliminating the hydrodynamic effect of water currents (Legendre & Troussellier, 1993).
- 3) Env (20x3): ASCII file containing three environmental variables: NH<sub>4</sub>, Phaeopigments, DNA(F).
- 4) Thau XY (20x2): ASCII file containing the geographic coordinates (X, Y) of the 20 sampling stations.

First, one should prepare the following distance matrices with program SIMIL. If a measure of similarity is used, instead of a distance, the corresponding file should be converted to distances using program CONVERT.

- 5) Esp/D1: Matrix of Euclidean distances [or any other distance of one's choice] among stations, computed from the species data (BNA and MA) of matrix (1).
- 6) Env/D1: Matrix of Euclidean distances among stations, computed from the environmental data of matrix (3).
- 7) Spatial/D1: Geographic (= Euclidean) distance matrix among stations, computed from the X and Y geographic coordinates of matrix (4).

## Autocorrelation program of Rv3; Autocor module of Rv4

The variable to be analysed is in matrix (2), called Chla (20x1); the data are quantitative, not nominal. The inter-point distances are to be found in the SIMIL matrix of geographic distances that you have computed, called Spatial/D1; it will be used to compute the distance classes. You will have to choose between equal-frequency or equal-distance classes, and to decide the number of classes that seems the most appropriate. Remember that for a set of  $n$  points, there are  $m = n(n-1)/2$  distances in the half distance matrix. Yule's empirical rule  $3.3 \times m^{0.25}$  may provide a first approximation of what an interesting number of distance classes might be.

When the program starts, the user has to provide a file name for the results of the calculations: division into classes and correlograms.

Ask the program to save the matrix of distances recoded into classes (the question actually is: "Save the data file recoded into classes?"); this becomes your file (8). You will need it to compute the Mantel correlogram (below).

## Simple Mantel test

Program MANTEL 3.0 (or module **Mantel** in Rv4) allows the comparison of distances in matrices (5) and (6), or (5) and (7), or else (6) and (7), using a linear statistic. The two matrices compared by the program are called A and B. Both matrices play the same role and may be declared indifferently as A or B. Request a "Mantel test for two matrices", and specifying that matrix A is a "Distance file from SIMIL". Since this is a small problem, you may request either 99, 499, or 999 permutations for this problem, or else 9999 if you are working on a fast machine. If a "normalized Mantel statistic" is requested, the program actually computes a Pearson product-moment linear correlation between the values in the two distance matrices.

["Similarity-Distance matrix rewritten?" — Answer "No". This is a remnant from a previous version that should disappear.]

*Note* – This program was developed under Macintosh System 6. Under System 7, the messages requesting the names of matrices A and B are not properly displayed. One has to know that for simple Mantel tests, the program first requests the name of matrix B, then that of matrix A. For partial Mantel tests, the program requests the name of B first, then C, and finally A. Use *The R Package v. 4* instead.

### **Partial Mantel test** (program **Mantel** of Rv3 or module **Mantel** of Rv4)

With MANTEL 3.0, it is important here to clearly declare which matrix plays the role of A, B and C. The program computes a Smouse, Long & Sokal partial Mantel statistic, noted AB•C, which is the partial linear correlation of A with B after removing the linear effect of C; partial correlations are often noted AB|C in textbooks of statistics. For this exercise, you may use all three distance matrices (5), (6) and (7) computed previously, in the order that you find appropriate. Use *The R Package v. 4* instead.

### **Mantel correlogram** (program **Mantel** of Rv3 or module **Mantel** of Rv4)

Here again, the order of the matrices is of utmost importance. After requesting a "Mantel test for two matrices", declare that matrix A is an "Input file in classes (Mantel correlogram)". When the identity of matrix A is requested, provide the name of the matrix of distance classes (8) that has been obtained when computing Moran's and Geary's correlograms. Matrix B is one of the other SIMIL distance matrices that you have computed: (5) or (6). In this example, the best correlogram is obtained using matrix (6).

Don't forget that the number of permutations that you request will be computed for each distance class in turn. With a slow machine, it may be unwise to request more than, say, 99 permutations in an exploratory analysis. Modern machines allow you to test using 999 or 9999 permutations.

The correlogram values are written down in the ASCII output file whose name has been provided by the user. As indicated by the program, the signs of the autocorrelation coefficients are reversed if matrix B is a *Distance* matrix. In other words, a negative sign indicates positive autocorrelation in that case, and vice versa. However, if matrix B is a *Similarity* matrix, then a positive sign indicated positive autocorrelation. Can you explain why?

### **References**

- Amanieu, M., P. Legendre, M. Troussellier & G.-F. Frisoni. 1989. Le programme Écothau: théorie écologique et base de la modélisation. *Oceanologica Acta* 12: 189-199.
- Legendre, P. & M. Troussellier. 1993. Origin of spatial structures in aquatic bacterial communities: From hypotheses to numerical solutions. Pp. 353-358 *in*: Guerrero, R. & C. Pedrós-Alió (eds.) *Trends in Microbial Ecology*. Proc. 6th Intern. Symp. Microbial Ecology (ISME-6), Barcelona, 6-11 September 1992. Spanish Society for Microbiology.
- Legendre, P., M. Troussellier, V. Jarry & M.-J. Fortin. 1989. Design for simultaneous sampling of ecological variables: from concepts to numerical solutions. *Oikos* 55: 30-42.