

On identifying significant co-occurrence of species in space and time

J. M. Vargas, R. Real & L. J. Palomo

Vargas, J. M., Real, R. & Palomo, L. J., 1997. On identifying significant co-occurrence of species in space and time. *Misc. Zool.*, 20.2: 49-58.

On identifying significant co-occurrence of species in space and time.— In the present paper the concepts "chorotype" and "chronotype" are defined operationally as the groups of species whose distributions overlap more than expected at random in space and time, respectively. A probabilistic method for testing the existence of such significant groups of species starting from a classification analysis of species according to either their geographical or temporal distribution is proposed. Two examples of the operational use of these concepts in biogeography are also provided. In example 1 we test for the existence of chronotypes of migrant soaring birds through the Straits of Gibraltar; in example 2 we not only detect chorotypes of micromammals in Europe, but also investigate the possible causes of the shared distributions.

Key words: Chorotype, Chronotype, Probabilistic classification, Logistic regression, Environmental factors, Biogeography.

(*Rebut: 28 VI 96; Acceptació condicional: 28 I 97; Acc. definitiva: 6 X 97*)

Vargas, J. M., Real, R. & Palomo, L. J., Depto. de Biología Animal, Fac. de Ciencias, Univ. de Málaga, 29071 Málaga, España (Spain).

This research has been partially supported by DGICYT, Project PB 93-1002.

Introduction

From a biogeographical point of view, a set of species may share distribution patterns in space or time or, on the contrary, may be distributed independently of each other. If the former is true, then the species are to be placed in a theoretically discontinuous sequence, either in space or time. Some biogeographical categories have been proposed that correspond to these discontinuous sequences, and which we will name "chorotypes" (BARONI-URBANI et al., 1978) for space, and "chronotypes" for time. The term "chronotype" can refer to migration patterns of birds, successional stages of vegetation, seasonal occupation of habitats, or even the groups of species belonging to the same geological period.

Defined in this way, each species constitutes a chorotype and a chronotype in itself, but the aim of the biogeographer is to detect some spatial or chronological patterns in the distributional data shared by several species. If such patterns exist, it is not necessary to invoke a different process to explain distribution of each species, but some factors may be responsible for the distributions shared by each group of species, and the overall environmental or historical interpretation may be more comprehensible. However, chorotypes and chronotypes do not refer to the ecological relationships between the species implied in the term guild or ecological community, because the species that share a geographical range may occupy different niches and even belong to different communities.

In the following, the meaning of the biogeographical categories "chorotype" and "chronotype" is discussed and a quantitative technique that allows us to test for their existence and define their limits displayed. Two examples of their application are also provided.

The coefficient model in association analysis

Similarity indices are frequently used to study the coexistence of species in an approach called the coefficient model (see JACKSON et al., 1992). However, SIMBERLOFF & CONNOR

(1979) stated that most indices of similarity are not associated with probability values, since their underlying distributions are unknown, thus preventing high and low levels of association between species from being recognized objectively with regard to what may be expected at random. The only distributions which have been studied in the literature are the simple matching coefficient (GOODALL, 1967), the Baroni-Urbani and Buser coefficient (BARONI-URBANI & BUSER, 1976), the Jaccard coefficient (BARONI-URBANI, 1980; REAL & VARGAS, 1996), and the Phi coefficient (JACKSON et al., 1992). Thus, these are the indices suitable for testing the statistical significance of similarity values.

Classification and ordination

When a matrix of similarity coefficients between species has been obtained, it may be analysed in two ways: by ordination, that is, by attempting to arrange the species within a theoretically continuous sequence; or by classification, the aim of which is to place the species in discontinuous groups (McCoy et al., 1986). In many studies both methods are used as complementary analyses (KENT & BALLARD, 1988). However, there is no objective and clear criterion for preferring one analysis to the other according to the data. A conservative approach to this problem may be to consider the existence of a continuous framework as the null hypothesis, and to test the groups obtained in a classification analysis against this null hypothesis.

Certainly, one cannot infer that a group structure exists from the partition of data resulting from a classification. The results of classification analyses are usually displayed as dendrograms. WHITTAKER (1975), BIRKS (1976), MCCOY et al. (1986) and BELBIN & McDONALD (1993), among others, pointed out that some distortion is caused by representing a multidimensional similarity matrix by a bidimensional dendrogram. This distortion can lead us to accept artificial groups of species that do not correspond to what is actually found in nature (KENT & BALLARD, 1988; VARGAS, 1993).

The method of McCoy et al. (1986), as modified by REAL et al. (1992) was used to detect significant chorotypes and chrono-

types. The main advantages of this method are that the necessary calculus is simple, and the significant groups are obtained on a probabilistic basis. So, the species are not considered as chorotypes or chronotypes when the relation between their distributions is compatible with that expected at random.

According to McCoy et al. (1986), the similarity values are to be transformed into three classes identified by "+", "-" and "0" signs, according to whether the values are significantly higher, lower or the same as expected at random. The method of McCoy et al. (1986) is modified to test for the existence of either strong or weak significant biogeographical segregations between the species.

For each dendrogram node a submatrix of significant similarities that only included the species involved in the node was established. This submatrix was divided into three zones: zone A and zone B, which corresponded to each group of species separated by the node; and zone AxB, corresponding to the intersection between both zones.

Pp(AxA) is the number of pluses between each pair of species within zone A divided by the total number of elements possible for comparisons in zone A. So, Pp(AxA) is the proportion of pluses in zone A. Psp(AxA) is the number of species in zone A that have at least one plus divided by the total number of species for comparisons in zone A, which represents the proportion of species in zone A with at least one plus. So its possible to compute d1(AxA) as follows: if the number of pluses in A is zero, then d1(AxA)=0; otherwise,

$$d1(AxA) = \frac{Pp(AxA) \times Psp(AxA)}{\sqrt{(Pp(AxA))^2 + (Psp(AxA))^2}}$$

Pm(AxA) and Psm(AxA) are defined as the proportion of minuses in zone A and the proportion of species in zone A with at least one minus, respectively, and these are computed in the same way as Pp(AxA) and Psp(AxA), but by taking into account the minuses. So d2(AxA) is defined in the following way: if the number of minuses in zone A is zero, then d2(AxA)=0; otherwise,

$$d2(AxA) = \frac{Pm(AxA) \times Psm(AxA)}{\sqrt{(Pm(AxA))^2 + (Psm(AxA))^2}}$$

Pp(AxB) and Psp(AxB) are defined in a similar way as Pp(AxA) and Psp(AxA), but referred to zone AxB. Then d4 is zero when the number of pluses in AxB is zero; otherwise,

$$d4 = \frac{Pp(AxB) \times Psp(AxB)}{\sqrt{(Pp(AxB))^2 + (Psp(AxB))^2}}$$

The parameters DW(AxA) and DW(BxB) measure to what extent the similarities higher than expected (+) tend to be in zones A and B, but not in AxB (see McCoy et al., 1986):

DW(AxA) = d1(AxA) - d2(AxA) - d4
analogously,

DW(BxB) = d1(BxB) - d2(BxB) - d4

where d1(BxB) and d2(BxB) are calculated as d1(AxA) and d2(AxA), but computing the pluses and minuses in zone B.

The parameter d3 is defined in the same way as d4, but computing the minuses in AxB. The parameter DS gives a measure of whether the similarities lower than expected (-) tend to be located in AxB, but not in A or B.

DS = d3 - d4 - d2(AxA) - d2(BxB).

The statistical significance of the node was assessed by a G test of independence of the distribution of the signs "+", "-" and "0" in the three zones of the submatrix, giving the parameters GW, for weak segregations, and GS, for strong segregations. If similarities higher than expected (+) tend significantly to be in zones A or B, but not in AxB, that is, if DW > 0 and GW is significant, then there is at least a weak segregation between both groups of OGUs, and the group whose DW is more than zero constitutes a chorotype or a chronotype. If similarities lower than expected (-) significantly tend to be located in AxB, but not in A or B, that is if DS > 0 and GS is significant, then a strong segregation exists between the groups.

Chorotypes and chronotypes

The result of this classification method can produce chorotypes or chronotypes, or the non-existence of significant groups. The existence of a significant group indicates that the ranges of the species included in the group overlap more than expected at random. If significant groups do exist, they may be strongly segregated or weakly segregated. A strong segregation signifies that the ranges of the groups overlap less than ex-

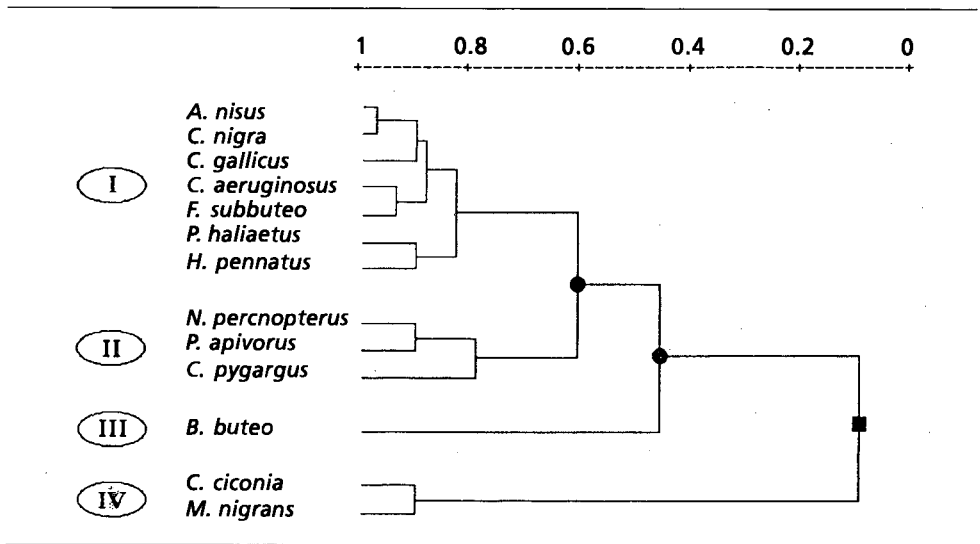


Fig. 1. Chronotypes (I - IV) established through the Straits of Gibraltar during the autumn migration of 15 soaring birds: ■ Strong segregation; ● Weak segregation.

Cronotipos establecidos durante el periodo migratorio postnupcial de 15 especies de aves planeadoras a través del Estrecho de Gibraltar (I-IV): ■ Segregación fuerte; ● Segregación débil.

where D is the number of weeks in which neither of the two species are present; A and B represent the number of weeks in which only species a and only species b are present, respectively; and C represents the number of weeks in which both species are recorded together. This index, like all those indices taking negative matches into account, changes its values when additional species whose migration periods include weeks in which the two species analysed are absent, are included in the analysis (BUSER & BARONI-URBANI, 1982).

UPGMA (Unweighted Pair-Group Method using arithmetic Averages) was used as the classification method, because it uses an agglomerative algorithm to carry out the classification and produces the least distortion of the original similarities in this type of algorithm (SNEATH & SOKAL, 1973). Agglomerative methods are preferred because as the null hypothesis of continuity implies that the species distribute independently of each other, it is the association between them that has to be demonstrated.

The matrix of significant similarities (table 1) was obtained using the table in BARONI-URBANI & BUSER (1976).

Four chronotypes (table 2; fig. 1) were obtained. *Ciconia ciconia* and *Milvus migrans* constituted an early chronotype strongly segregated from the rest of species. The three other chronotypes are segregated weakly. The existence of chronotypes indicates that these bird species do not migrate independently from each other, but that some of them associate in shared periods of migration which may be statistically distinguished from other periods of migration shared by other species. *Buteo buteo* constitutes a chronotype formed by one species only, indicating that this species does not associate with any other in its chronological pattern of migration.

The chronological pattern defined by the four chronotypes facilitates the searching of the causes of the periods of migration. The causes of the particular period in which each bird species migrate are difficult to assess, because the researcher must investigate 13 different causes related to the conditions

Table 2. Significant segregation between the species groups established in the dendrogram. GW and GS show weak and strong segregation between respective groups: * $p < 0.005$; ** $p < 0.001$; *** $p < 0.005$. DW (AxA) and DW (BxB) show the homogeneity inside each group analysed.

*Segregaciones significativas entre los grupos de especies establecidos en el dendrograma. GW y GS indican, la segregación débil y fuerte entre los respectivos grupos: * $p < 0,005$; ** $p < 0,001$; *** $p < 0,005$. DW (AxA) y DW (BxB) son los parámetros que estiman la homogeneidad interna de cada grupo.*

Groups combined by UPGMA			Segregation					
			Agregation		Weak		Strong	
Group A	Group B	Coefficient	DW(AxA)	DW(BxB)	GW	p	GS	p
9-7	4-3	0.6295	0.5291	0.5291	33.4055	***	0.0	N.S.
9-3	13	0.4960	0.4407	-0.0877	7.6517	**	0.3109	N.S.
9-13	1-2	0.1284	0.4483	0.7071	21.0113	***	31.4916	***

of the breeding areas, the wintering areas, the physiology of the organisms or their feeding habits. However, it is more easy to search for the causes of four chronotypes of migrant species than for the causes of the migrant periods of the 13 individual species.

Example 2: Chorotypes of micromammals (insectivores and rodents) in Europe and the environmental factors related to them

Europe was divided into 53 major physiological regions and recorded the presence of 24 insectivore and 62 rodent species in each. The distribution maps of NIETHAMMER & KRAPP (1978, 1982, 1990), updated with unpublished data from the Iberian penin-

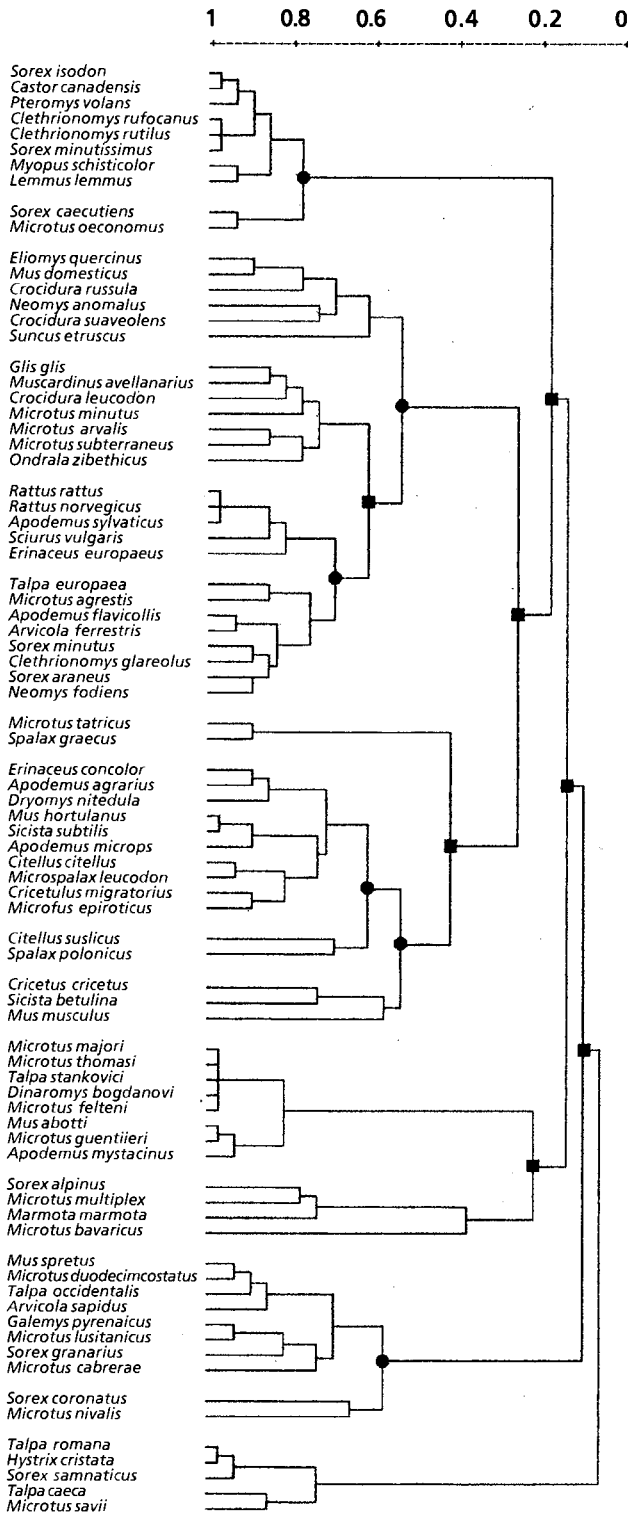
sula compiled by the Spanish Mammal Society (SECEM) and different national distribution atlases (PUCEK & RACZYNSKI, 1983; FAYARD, 1984; ARNOLD, 1993) were used.

The Baroni-Urbani and Buser's index and UPGMA, followed by the probabilistic method described above was applied to this matrix of presences.

Fifteen chorotypes were obtained (see fig. 2) on which we carried out logistic regressions using the following environmental variables: annual range of temperature (RT), mean annual temperature (T), mean temperature of January (TJA), mean temperature of July (TJU), potential evapotranspiration (PET), annual mean precipitation (P), actual evapotranspiration (AET), and distance to the Urals (DUR). P, T, TJA, TJU

Fig. 2. European chorotypes of micromammals (insectivores and rodents). Species that do not associate with others have been excluded from the dendrogram: ■ Strong segregation; ● Weak segregation.

Corotipos europeos de micromamíferos (insectívoros y roedores). Se han excluido del dendrograma las especies que no se asocian significativamente con ninguna otra especie: ■ Segregación fuerte; ● Segregación débil.



and RT values were obtained from the UNESCO (1970). PET and AET were derived from the maps of the USSR NATIONAL COMMITTEE FOR THE INTERNATIONAL HYDROLOGICAL DECADE (1977). The distance from the Urals (DUR) was estimated using THE TIMES ATLAS OF THE WORLD (1988).

A logistic regression (SHOTT, 1991) was carried out to characterize the distribution area of each chorotype environmentally (see, for example, ROMERO & REAL, 1996). With this analysis the following equation was obtained:

$$p = e^y / 1 + e^y$$

where "e" is the base of the Napierian logarithm, "p" is the probability that the chorotype is present in the physiographical region and "y" is a regression equation of the form:

$$y = a + b x_1 + c x_2 + \dots + n x_n$$

where x_i are a selection of the environmental variables used.

The previous equation was used to estimate the logarithm of the odds of the pres-

ence of the chorotypes. The odds of an event is the probability of the event occurring divided by the probability of this event not occurring. So, the odds of finding a chorotype in a particular region is given by the logistic regression analysis, which also selects the environmental parameters that determine these events.

The logistic model was tested by the chi-square test of goodness of fit. The estimation of the parameters a, b, etc. in the previous equation was by maximum likelihood and was tested by the test of Wald.

Using logistic regression six possible causal factors for the chorotypes were tested:

1. Climatic stability. The zones with the greatest climatic stability favour those groups of species that have strict eco-physiological requirements. Stability of temperature by the differences between the mean temperatures of January and July (RT) were measured.

2. Energy availability. The presence of the

Table 3. Chorotypes of micromammals in Europe correctly explained by logistic regressions according to the environmental factors analysed. The percentage expresses the physiographical regions correctly classified as either belonging or not to the distribution area of each chorotype.

Corotipos de micromamíferos en Europa correctamente explicados por regresión logística en relación a los factores ambientales analizados y porcentaje de regiones fisiográficas correctamente clasificadas en función de su pertenencia o no al área de distribución de cada corotipo.

Chorotype	Environmental factor	Equation	% correct
II	Penetration route	$y = -0.0037 \text{ DUR} + 8.42$	94.34
III	Energy availability	$y = 0.0210 \text{ PET} - 12.31$	94.34
VI	Climatic stress	$y = -1.3700 \text{ TJU} + 32.54$	96.23
VIII	Penetration route	$y = -0.0024 \text{ DUR} + 6.44$	83.02
X	Climatic stress	$y = -0.0078 \text{ PET} + 6.68$	81.13
XII	Simultaneous availability of water and energy	$y = 0.0580 \text{ AET} - 35.16$	92.45
XIV	Simultaneous availability of water and energy	$y = 0.0120 \text{ AET} - 7.00$	73.58

chorotype may be limited by the available environmental energy. The variables used to measure available energy are the mean annual temperature (T), the potential evapotranspiration (PET) and the mean January temperature (TJA).

3. Water availability. Distributions of some species are controlled by the availability of water. The variable used here to measure water availability is the mean annual precipitation (P).

4. Simultaneous availability of water and energy. Water and energy are biologically useful when they are available simultaneously (STEPHENSON, 1990). The variable that measures this factor is the actual evapotranspiration (AET) (ROSENZWEIG, 1968).

5. Climatic stress. Extreme environmental conditions may subject species to climatic stress that very often goes beyond their levels of tolerance. Increases in the potential evapotranspiration (PET), mean annual temperature (T), and mean July temperature (TJU) can restrict the distributions of species sensitive to climatic stress.

6. Penetration routes of biotas. The number of species is inversely related to the geographic distance from the zones of penetration of biotas. The variable related to this factor is the distance from the Ural Mountains (DUR).

Seven chorotypes have been explained according to the six preceding causal factors (table 3). Energy availability explains one of the chorotypes; simultaneous availability of water and energy explains two more; climatic stress explains another two chorotypes; and penetration routes explains the other two.

Eight chorotypes are not explained with the factors considered a priori, although some of them may be associated to the Mediterranean peninsulas, where the species probably differentiated.

Acknowledgements

We thank Dr. Eusebio García-Manrique, Dra. M^a Luisa Gómez, J. Carlos Guerrero, Dra. Ana L. Márquez, Jesús Olivero and Ana B. Rojas for their helpful collaboration, comments and suggestions about the manuscript.

Resumen

Identificación de solapamientos significativos de especies en el espacio y en el tiempo

En el presente trabajo se define los conceptos de "corotipo" y "cronotipo" de forma operativa, es decir, como grupos de especies cuyas distribuciones se solapan respectivamente en el tiempo y en el espacio más de lo esperado por el azar, sin que ello implique relación ecológica directa entre las especies. Se propone un método probabilístico para poner a prueba la existencia de tales grupos significativos de especies, partiendo de un análisis de clasificación de las especies consideradas en función de sus distribuciones geográficas o temporales. Se aportan dos ejemplos en relación al empleo de los citados conceptos en biogeografía. En el ejemplo 1 se investiga la existencia de cronotipos de especies planeadoras migrantes a través del Estrecho de Gibraltar; se obtienen cuatro cronotipos diferentes (tabla 1, tabla 2, fig. 1) que corresponden a grupos de especies cuyos periodos de migración difieren estadísticamente entre sí. En el ejemplo 2 además de establecer los posibles corotipos europeos de micromamíferos (fig. 2), se investiga las causas aparentes que determinan su configuración, utilizando regresiones logísticas para caracterizar, desde un punto de vista ambiental, el área de distribución de cada corotipo (tabla 3).

References

- ARNOLD, H. R., 1993. *Atlas of mammals in Britain*. Natural Environment Research Council, London.
- BARONI-URBANI, C., 1980. A statistical table for the degree of coexistence between two species. *Oecologia*, 44: 287-289.
- BARONI-URBANI, C. & BUSER, M. W., 1976. Similarity of binary data. *Syst. Zool.*, 25: 251-259.
- BARONI-URBANI, C., RUFFO, S. & VIGNA TAGLIANTI, A., 1978. Materiali per una biogeografia italiana fondata su alcuni generi di coleotteri cicindelidi, carabidi e crisomelidi. *Estr. Mem. Soc. Ent. Ital.*, 56: 35-92.
- BELBIN, L. & McDONALD, C., 1993. Comparing three classification strategies for use in ecology. *J. Veg. Sci.*, 4: 341-348.

- BERNIS, F., 1980. *La migración de las aves en el Estrecho de Gibraltar*. Vol. I. Pub. Cát. Vert., Univ. Complutense Madrid, Madrid.
- BIRKS, H. J. B., 1976. The distribution of Pteridophytes: a numerical analysis. *New Phytol.*, 77: 257-287.
- BUSER, M. W. & BARONI-URBANI, C., 1982. A direct nondimensional clustering method for binary data. *Biometrics*, 38: 351-360.
- CONNOR, E. F. & SIMBERLOFF, D., 1983. Intraspecific competition and species co-occurrence patterns on islands: Null models and the evaluation of evidence. *Oikos*, 41: 455-465.
- FAYARD, A. (Ed.), 1984. *Atlas des Mammifères sauvages de France*. Société Française pour l'Etude et la Protection des Mammifères, Paris.
- GOODALL, D. W., 1967. The distribution of the matching coefficient. *Biometrics*, 23: 647-656.
- JACKSON, D. A., SOMERS, K. M. & HARVEY, H. H., 1992. Null models and fish communities: Evidence of nonrandom patterns. *Am. Nat.*, 139: 930-951.
- KENT, M. & BALLARD, J., 1988. Trends and problems in the application of classification and ordination methods in plant ecology. *Vegetatio*, 78: 109-124.
- MCCOY, E. D., BELL, S. S. & WALTERS, K., 1986. Identifying biotic boundaries along environmental gradients. *Ecology*, 67: 749-759.
- NIETHAMMER, J. & KRAPP, F., 1978. *Handbuch der Säugetiere Europas*. Band 1 (Nagetiere I). AULA-Verlag, Wiesbaden.
- 1982. *Handbuch der Säugetiere Europas*. Band 2/1 (Nagetiere II). AULA-Verlag, Wiesbaden.
- 1990. *Handbuch der Säugetiere Europas*. Band 3/1 (Insektenfresser, Primaten). AULA-Verlag, Wiesbaden.
- PUCEK, Z. & RACZYNSKI, J., (Eds.), 1983. *Atlas of Polish mammals maps*. Polish Scientific Publisher, Warszawa.
- REAL, R. & VARGAS, J. M., 1996. The probabilistic basis of Jaccard's index of similarity. *Syst. Biol.*, 45: 380-385.
- REAL, R., GUERRERO, J. C. & RAMÍREZ, J. M., 1992. Identificación de fronteras bióticas significativas para los anfibios en la cuenca hidrográfica del Sur de España. *Doñana, Acta Vertebrata*, 19: 53-70.
- ROMERO, J. & REAL, R., 1996. Macroenvironmental factors as ultimate determinants of distribution of common toad and natterjack toad in the South of Spain. *Ecography*, 19: 305-312.
- ROSENZWEIG, M. L., 1968. Net primary productivity of terrestrial communities: prediction from climatological data. *Am. Nat.*, 102: 67-74.
- SHOTT, S., 1991. Logistic regression and discriminant analysis. *JAVMA*, 198: 1902-1905.
- SIMBERLOFF, D. & CONNOR, E. F., 1979. Q-mode and R-mode analyses of biogeographic distributions: Null hypotheses based on random colonization. In: *Contemporary quantitative ecology and related econometrics*: 123-138 (G. P. Patil & M. Rosenzweig, Eds.). International Cooperative, Fairland, Maryland.
- SNEATH, P. H. A. & SOKAL, R. R., 1973. *Numerical taxonomy. The principles and practices of numerical classification*. Freeman, San Francisco.
- STEPHENSON, N. L., 1990. Climatic control of vegetation distribution: the role of the water balance. *Am. Nat.*, 135: 649-670.
- THE TIMES ATLAS OF THE WORLD, 1988. Times Books, London.
- UNESCO, 1970. *Climatic Atlas of Europe*. UNESCO, Paris.
- USSR NATIONAL COMMITTEE FOR THE INTERNATIONAL HYDROLOGICAL DECADE, 1977. *Atlas of world water balance*. UNESCO, Paris.
- VARGAS, J. M., 1993. Siete pecados capitales en Biogeografía. *Zool. Baetica*, 4: 39-56.
- WHITTAKER, R. H., 1975. *Communities and ecosystems*. Macmillan, New York.